

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 13:18:50 ; Search time 41 seconds
(without alignments)
868.274 Million cell updates/sec

Title: US-09-494-297-2
Perfect score: 3945
Sequence: 1 MKKTRPPNKNITNTQRLVLS.....IAGISLGIMGIIIRIRIKHD 757

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
1	148	3.8	1183	1	CNA_STPAU
2	139	3.5	621	1	HTPG_HELPY
3	139	3.5	1433	1	SUBF_BACSU
4	139	3.5	2366	1	TOXB_CLODI
5	137	3.5	1177	1	V307_MYCGE
6	136.5	3.5	4590	1	FATH_HUMAN
7	135	3.4	1176	1	SLAP_BACSH
8	135	3.4	1228	1	SLAP_BACST
9	135	3.4	1301	1	SAC3_YEAST
10	132.5	3.4	1116	1	SLPH_BRECH
11	132.5	3.4	3097	1	CADN_DROME
12	130.5	3.3	881	1	LHSL_YEAST
13	130.5	3.3	908	1	DPOI_BORBU
14	130.5	3.3	1020	1	RPGR_HUMAN
15	130	3.3	621	1	HTPG_HELPY
16	130	3.3	1372	1	FUSL_SCHPO
17	130	3.3	3381	1	PGCV_BOVIN
18	129	3.3	461	1	P55G_HUMAN
19	129	3.3	1292	1	RPQC_MYCGE
20	128.5	3.3	914	1	PBPB_BACSU
21	128.5	3.3	1182	1	RPQC_CLOAB
22	128.5	3.3	1828	1	MAP2_MOUSE
23	128	3.2	1394	1	HAP_HAEIN
24	127	3.2	1628	1	NAGH_CLOPE
25	127	3.2	1650	1	VIT6_CABEL
26	126.5	3.2	822	1	DEXT_STEEL
27	126.5	3.2	1772	1	MSPI_PLAYO
28	126	3.2	1036	1	V414_MYCGE
29	126	3.2	1409	1	HAPL_HAEIN
30	125	3.2	700	1	HS9C_DICDI
31	125	3.2	862	1	SLA2_BACAN
32	124.5	3.2	461	1	P55G_BOVIN
33	124.5	3.2	1251	1	RBP2_PLAVB

34	124.5	3.2	3060	1	BPER_HUMAN
35	124.5	3.2	6359	1	BACC_BACLI
36	124	3.1	2334	1	WAPA_BACSU
37	123.5	3.1	1053	1	SLPM_BACBR
38	123.5	3.1	1358	1	SIR4_YEAST
39	123.5	3.1	3063	1	CA1C_HUMAN
40	123	3.1	861	1	P55G_MOUSE
41	123	3.1	831	1	ION_HELPY
42	123	3.1	980	1	BOB1_YEAST
43	123	3.1	1022	1	SCA4_RICPR
44	123	3.1	1055	1	CC28_SCHPO
45	122.5	3.1	681	1	BRC2_HALRO

ALIGNMENTS

RESULT 1

ID	CNA_STPAU	STANDARD	PRT: 1183 AA.
AC	Q53654		
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Collagen adhesin precursor.		
GN	CNA.		
OS	Staphylococcus aureus.		
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.		
OX	NCBI_TaxID=1280;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FDA 574;		
RX	MEDLINE=92165839; PubMed=1311320.		
RA	Patti J.M., Jonsson H., Guss B., Switalski L.M., Wilberg K.,		
RT	Lindberg M., Hoeoek M.;		
RT	"Molecular characterization and expression of a gene encoding a		
RL	staphylococcus aureus collagen adhesin.";		
RL	J. Biol. Chem. 267:4766-4772(1992).		
RN	[2]		
RP	ERRATUM.		
RA	Patti J.M., Jonsson H., Guss B., Switalski L.M., Wilberg K.,		
RA	Lindberg M., Hoeoek M.;		
RL	J. Biol. Chem. 269:11672-11672(1994).		
RN	[3]		
RP	COLLAGEN-BINDING DOMAIN.		
RC	STRAIN=FDA 574;		
RX	MEDLINE=94032261; PubMed=8218209;		
RA	Patti J.M., Boles J.O., Hoeoek M.;		
RT	"Identification and biochemical characterization of the ligand		
RT	binding domain of the collagen adhesin from staphylococcus aureus.";		
RL	Biochemistry 32:11428-11435(1993).		
RN	[4]		
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.		
RX	MEDLINE=97475225; PubMed=9334749;		
RA	Smeyers-Verbeke J., Patti J.M., Carson M., House-Pompeo K., Teale M.,		
RA	Moore D., Jin L., Schneider A., Delucas L.J., Hoeoek M.,		
RA	Narayana S.V.L.;		
RT	"Structure of the collagen-binding domain from a Staphylococcus		
RT	aureus adhesin.";		
RL	Nat. Struct. Biol. 4:833-838(1997).		
CC	-1- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO		
CC	COLLAGEN-CONTAINING SUBSTRATA.		
CC	-1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by		
CC	an amide bond (potential).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		

```

DR EMBL: M61736; AAA20874.1; -.
DR PDB: 1AMX; 24-JUN-98.
DR PDB: 1D2O; 27-SEP-00.
DR PDB: 1D2P; 27-SEP-00.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR006192; LPXTG.
DR TIGRGRAMS: TIGR01167; LPXTG_anchor; 1.
DR PROSITE: PS00847; GRAM_POS_ANCHORING; FALSE_NEG.
DR Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.
KW SIGNAL.
FT CHAIN 1 29
FT PROPEP 30 1154
FT DOMAIN 151 318
FT DOMAIN 533 1093
FT DOMAIN 1093 1157
FT REPEAT 533 719
FT REPEAT 720 906
FT REPEAT 907 1093
FT SITE 1151 1155
FT MOD_RES 1154 1154
FT STRAND 174 179
FT TURN 182 183
FT STRAND 185 186
FT STRAND 187 194
FT TURN 196 197
FT STRAND 201 201
FT STRAND 205 211
FT STRAND 215 228
FT TURN 229 230
FT STRAND 232 234
FT STRAND 239 246
FT HELIX 248 249
FT STRAND 251 255
FT TURN 256 259
FT STRAND 260 265
FT HELIX 267 270
FT TURN 271 272
FT STRAND 273 283
FT TURN 286 287
FT STRAND 290 299
FT STRAND 301 301
FT TURN 302 303
FT STRAND 307 311
FT STRAND 314 317
SQ SEQUENCE 1183 AA; 133066 MW; B6A1CC072E575D76 CRC64;

Query Match      3.8%; Score 148; DB 1; Length 1183;
Best Local Similarity 19.7%; Pred. NO. 1;
Matches 165; Conservative 123; Mismatches 299; Indels 252; Gaps 44;

OY 64 PSSSEYRWY---GYESYVRGHPYK-----QFRVADLRVNLGSRSY---QVYCF 109
   | : : | : : | : : | : : | : : | : : | : : | : : | : : |
DB 182 PEDTTHVRWFLNINNKSYVKDITIKDIOGGQGLDSLNLNINVTGTHSNYSGSAIT 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 110 NLKKAFFLS-----DSSVKKWKTKHDGISTKREDYAMSPRTG----- 148
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 242 DEKAPR-OSKITVDNTKNTIDVTIPQGGVSYNSFSINK---TKTINEQKEFVNNS 295
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 149 -----DELNOKLRVWYNGHPONANGIMG-----LEPLNALRVQEAIVWYSDNA 194
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 296 QAWYOHGKEEVNGK--SFNHVYHNINANAGIDGYKGLKYLKQKQDKT-----A 344
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 195 PISNPDESFK-RESESNLVSTQSLMRQALQDLDPNLATKPKQVPPDFOLSIPESE- 252
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 345 PLAN--VKFKLSKSDSVKNDQ---KEIEITDANGIANIKALPSGDYILKEIAPR 397
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 253 ----DKGDKY-----NKGYONLISGL-----VPTKPTPEDPPMPNPQPTSY 293
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 398 PYTFPDKDEYPTFMKDTDNOGYFTTLENKALEKTKDVSQAOKWSESTOKYKP-----TI 451
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 294 LIRKVAIGYS-----KLLEGATLQLTGDVNNNSQARVFSSND--IGERIELSDGT 342
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 452 YKKLYKODDNQNTPTPVDAKAEIKKLEDGTTKVTWSNL-----PENDNKGAIK----- 498

```

```

OY 343 YLTLELN-----SPAGYSIAE-----PI-TFKVAGKVYTIIDGKOIENKEL- 385
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 499 YLKEVNAAGEDTTPREGYTKENGGLVNTNTERKLEFTTSISGEKWDKQKRPKEK- 556
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 386 VEPYSVAVNDPEFSPVLTQNTAKRY-----AKKNSSQVYVCENADLKSPSESG 440
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 557 ----VSYNLLANGERKVTLDVTSFTNMKYEFKDLPKYDEG-KLLEYVTEDHVADYTDIN 612
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 441 GKTMDPEFTGEVKYTHIAGRDLEKTYVKKRDPDPTPLKIKVIEK----- 488
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 613 GTITTKYTTGETSAIVTKMWDNNODGKRPFIEIKVELYQDGAIKGKALINESNNWH 672
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 489 ----GYREKQALIEYSGLETQLRATQ-----LAIFYETDSALDKRLKYD 532
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 673 TWTGLDEKAKGQGVKQVTEELTVKGYTTHVDNNDGNLVTNKKYPTFTSISGEKWD- 721
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 533 HGFQDNDSTLVAKLIVEAODSNPQTLDDFFTPNNKKYQSL---IGTQHPEDVD 589
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 732 ----DKDN-----QDGKREKVSVN-LDADGERKVTLDVTSFTNMKYE-FKD 772
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 590 IIRMEDEKREIYPTVHNLTKRYTGLAGDRKDFHEIE-----LKNKQELLSQTYKT- 643
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 773 LPKYDESKTI-----EYIVT---EDHVKDYTTDINGTITTKYTPGETSAIVTKW 820
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 644 -DKTNLE-----FKDGRAT-----INLKHSGLTLAGLPE-----GYSYLVE- 680
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 821 DDNNQDGRKRPTEIKVELYQDGAIKGKATGKALINESNNWHWTWGLDEKAKGQVYVEEL 880
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 681 TDESGYKVKVYNSQEVANATYK-----IGITSEDLAEENKPEVYPPGYVDOKT--NG 731
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 881 TKVKGYYTHVDNNDGNLVTNKKYPTFTSISGEKWDKQKRPKEKVSYNLLANG 939
   : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 2
HTPG_HELPY
ID HTPG_HELPY STANDARD; PRT; 621 AA.
AC P56116;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chapterone protein htpg (heat shock protein htpg) (high temperature
DE protein G).
GN HTPG OR HP0210.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteriaceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervilange A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Meldman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -!- FUNCTION: Molecular chaperone. Has ATPase activity (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
CC or send an email to license@isb-sdb.ch).

DR EMBL: AE000541; AAD07278.1; -
DR PIR: B64546; B64546.
DR HSSP: P07900; 1YER.
DR TIGR: HP0210.
DR HAMAP: MF_00505; -; 1.
DR InterPro: IPR003594; ATPbind_ATPase.
DR InterPro: IPR001404; Hsp90.
DR Pfam: PF02518; HATPase_c; 1.
DR Pfam: PF00183; HSP90_1.
DR PRINTS: PR00775; HEATSHOCK90.
DR SMART: SM00387; HATPase_c; 1.
DR PROSITE: PS00298; HSP90; 1.
KW Chaperone; ATP-binding; Heat shock; Complete proteome.
FT DOMAIN 1 341 A; SUBSTRATE-BINDING (BY SIMILARITY).
FT DOMAIN 342 547 B (BY SIMILARITY).
FT DOMAIN 548 621 C; 28F198CIDC7EAB9A CRC64.
SQ SEQUENCE 621 AA; 71274 MW; 28F198CIDC7EAB9A CRC64.

Query Match 3.5%; Score 139; DB 1; Length 621;
Best Local Similarity 21.9%; Pred. No. 1.2;
Matches 130; Conservative 85; Mismatches 227; Indels 152; Gaps 30.

OY 201 ESFRRESSENN---LVSTSQLSLMRQALKQLDPLNLTAKMPQVDPDFQSLFSEEDKG- 255
DB 28 EIFRELVSNSADLKLNYLMLTDEKLK-----GLNTPPSIHLFSQCKTLTITNGMI 82
OY 256 --DKVN-----KGYONLSGGLVPTKPTPTGDPMPDPNQTSTVILRRKYAIDY 303
DB 83 GMDKNDLIEHGTAKSTKMFLSA-----LSGD-----KKKSALIGCGVGFY 127
OY 304 SKLEGATLQLTGDNVNSFOARVSSN-----DIGERIELSDGYITTELNSPAGY---- 354
DB 128 SAFVASKIVQTKRVNSDAQYAWMSDGKFEISECVKDSOGFEITFLDEDESHFASR 187
OY 355 -----SIAPETFKVEAKVYTIIDGK-QIENPKNEIPEYISVEAYNDFEER----- 400
DB 188 WEIDSVKKYSEHLPFPI-----FLTYDTTKHEGEGDNCKEIKKEKCEOINQASALMKMK 243
OY 401 SVLTTONYAKFYVKKNNKSSQVYVCFNADLKSPDSDGCKTMDPTTGFEVYKTHI-- 458
DB 244 SELDKDKYEFYQSFANDNSPELSTYHNK-----VEGSLTYTTTFY 284
OY 459 ---AGRDLFK-----YTVKPRDTPD---TFELKHKKVLEKG-----YREKG 494
DB 285 IPSTAPEDMFPRVDYKSGVKLYKRVFTDDDKELLPYLRFVKGVIDSEDLPLVNSRE-- 342
OY 495 QAIYESGLTQLRRAATQALAYFTDSAELEKDKLKHGCGMDSTLAAKILVE--Y 552
DB 343 --ILQOKIILANISASVYKI--LSEIERLSKDE-KNYKFEY-----PRGKYLEGLY 391
OY 553 AQDSNPOLTDLDFEIPNNKKYOSLIGTQWHPEDLVDIRREDKKEVIPV--HNLTLEKT 611
DB 392 GDFENKEKLELELRFYSKD--EKLSLKEKLENL-----KENQSTIYLLGENDLKA 444
OY 612 VTGLAGRTKDFHEIELKNNKOE--LLSQTVKTDKTNLEKRGKATINLKHGESLLOG 669
DB 445 SPL--EKYAKGVDVLLSDDEIDAFVMPGVNEDKT--PRDA-----SHSLSLKELG 494
OY 670 LPE-----GYSLVKTETDSEGYKVKVNSQEVANAYVSKITGITSDETLAEEN 716
DB 495 LEEIHDEVKQFDLMKAFE--ENLKEIKGVELSHLTSVAVALIGDEONAMMAN 547

RESULT 3
SUBF_BACSU STANDARD: PRT: 1433 AA.
AC P16397;
DT 01-AUG-1990 (Rel. 15, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bacillopeptidase F precursor (EC 3.4.21.-) (Esterase) (RP-I protease)
DE (90 kDa serine proteinase).
GN BPR OR BPF.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 195-222.
RX MEDLINE=90216713; PubMed=2106512;
RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Thierault K.A.,
RA Pero J.;
RT "Bacillopeptidase F of Bacillus subtilis: purification of the protein
RT and cloning of the gene";
RL J. Bacteriol. 172:1470-1477(1990).
RN [2]
RP REVISIONS.
RC STRAIN=168;
RX MEDLINE=90368623; PubMed=2118514;
RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Thierault K.A.,
RA Pero J.;
RL J. Bacteriol. 172:5520-5521(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90216713; PubMed=2108961;
RA Wu X.-C., Nathoo S., Pang A.S.-H., Carne T., Wang S.-L.;
RT "Cloning, genetic organization, and characterization of a structural
RT gene encoding bacillopeptidase F from Bacillus subtilis";
RL J. Biol. Chem. 265:6845-6850(1990).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96044033; PubMed=9384377;
RA Kunst F., Ogatawa N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Campano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Dentzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Guim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Jorjans B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Forwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolet C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Yamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weltzenegger T.,
RA Winiers P., Wipat A., Yamamoto H., Yaman K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumslein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis";
RL Nature 390:249-256(1997).
RN [5]
RP SEQUENCE OF 1-211 FROM N.A.
RX MEDLINE=89008108; PubMed=3139638;
RA Beall B., Lowe M., Lutkenhaus J.;
RT "Cloning and characterization of Bacillus subtilis homologs of
RT Escherichia coli cell division genesftsZ andftsA";
RL J. Bacteriol. 170:4855-4864(1988).
RN [6]

RP SEQUENCE OF 1410-1433 FROM N.A.
 RC STRAIN-168 / Marburg;
 RX MEDLINE=50174995; PubMed=2106671;
 RA Masuda E.S., Anaguchi H., Sato T., Takeuchi M., Kobayashi Y.;
 RT "Nucleotide sequence of the sporulation gene spoIIIGA from *Bacillus*
 RT subtilis";
 RL Nucleic Acids Res. 18:657-657(1990).
 RN [7]
 RP SEQUENCE OF 195-219.
 RC STRAIN-Natto 16;
 RA Kato T., Yamagata Y., Arai T., Ichishima E.;
 RT "Purification of a new extracellular 90-kDa serine proteinase with
 RT isoelectric point of 3.9 from *Bacillus subtilis* (natto) and
 RT elucidation of its distinct mode of action";
 RL Biosci. Biotechnol. Biochem. 56:1166-1168(1992).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: Belongs to peptidase family S8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M29035; AAA62679.1; -
 DR EMBL: J05400; AAA83362.1; -
 DR EMBL: Z99111; CAB13403.1; -
 DR EMBL: Z99112; CAB13404.1; -
 DR EMBL: M29630; AAA23458.1; -
 DR EMBL: X17344; CAA35224.1; -
 DR PIR: A36734; A36734.
 DR HSP: P00782; 2SMT.
 DR MEROPS: S08.017; -
 DR Subtilist: BG10233; hpr.
 DR Interpro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 KM Hydrolyase; Serine protease; Zymogen; Signal; Complete proteome.
 FT SIGNAL 1 30
 FT PROPEP 31 194
 FT CHAIN 195 755 POTENTIAL.
 FT PROPEP 756 1433 BACILOPEPTIDASE F.
 FT ACT_SITE 227 227 POTENTIAL.
 FT ACT_SITE 274 274 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 452 452 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CONFLICT 219 219 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CONFLICT 393 393 T -> A (IN REF. 7).
 FT CONFLICT 829 834 A -> V (IN REF. 3).
 FT CONFLICT 836 841 KHONKA -> N (IN REF. 3).
 FT CONFLICT 844 852 OPOVLV -> RTRLYS (IN REF. 3).
 FT CONFLICT 853 1433 AOVSVETG -> ECRGRHSV (IN REF. 3).
 FT CONFLICT 853 1433 MISSING (IN REF. 3).
 SQ SEQUENCE 1433 AA; 154577 MW; 98DF6846897807C9 CRC64;
 Query Match 3.5%; Score 139; DB 1; Length 1433;
 Best Local Similarity 20.7%; Pred. No. 4.3;
 Matches 165; Conservative 101; Mismatches 273; Indels 258; Gaps 43;
 QY 64 PDSSEY-----RWYGE-----SYV-----RGHYTKQFRAN-----DLRYNLEGRS 103
 DB 701 PDSGSLFLOKSNHNLNEDDFDYGVVLPBGEKNWEDAGVNGKTSWTDEIDLNAVKG 760
 QY 104 YQVYCNLKKAPLPGSSVYK--WYKKHHDGISTKFEVDYMSPRITGDEL----- 151
 DB 761 Q-----NIQMFVLDQDESTAKSGW-----IDDVVLSKDSACKTVKKKGLGYEK 806
 QY 152 --NOKLRVYVNGHPNAN-----GIMEGLEPLNA-IRYQOEAVMYSDNA----- 194
 DB 807 SGKQKKRVNPKAKKASANTAVKHNKALIQVLPRLKAOVSVYETGKSTSDSGTYL 866

QY 195 -----PISNDESEFKRESSESNLYS-----TSQSLMRQALQK-----LIDPNL---ATK 235
 DB 867 KHKAGDTLTAEAENVGYSKQOVSLKTDQTTQANFTLEEMKGTGLKGTIVINKTTGEPVYG 926
 QY 236 MPKQVPPDPQLSTFSESDDKDKYKQNLISGLVFTKPTTGPDPMPNPQOTTSVLI 235
 DB 927 ASYVVEDAAVEPAMTNDKGEYMLEAVE-----GAYTIKAAPR----- 965
 QY 296 RKYALGDYSKLECATLQ-----TGDNVSFOARFESSNDIGERIE 337
 DB 966 --YSDSEFVELKGDVTKETALRPFVGPGEIAYDGTAEANASYFA--AGNMAVKMT 1020
 QY 338 LSDG-----TYTLELNSPAGYSIAEPTEFEVAGKY--TIIDKOIENPKE 384
 DB 1021 LADGKDKMGLTGILFREMDEFPDPG-----TEFKVE--YDATGKDKGA---PGK 1067
 QY 385 IVEPYSVEAVN-----DPEEFSVLTONYAKFY-----YAKKNSSQVYVC 426
 DB 1068 IAGFNEALRNGEMTKVDLSSKGINVDKDFYLYIQSKDPDYPSPGLAMDETQNSGRW 1127
 QY 427 FNADLK-SPPSEDDG-----KTPEDFTTGEVK-YTHIAGRDLFEYTVKPRDTP 475
 DB 1128 QYIDGKMQPDDKADGNMIMALVDYEAAYEILSPDKSYTN--KD--SVYK-GNAP 1181
 QY 476 DFLK-----HIKKYIEKGYRE-KGAIEYSGILTEQLRAAQOLAI 515
 DB 1182 GTTVHIYNGEKEAGETKAADGTFHAGIILINKSENELTATASDNGTTD---ASSPIV 1237
 QY 516 YVFTDSAEILKDKLKVNGHFGDMNDSLVAKTLVEAODSNPOLDLDFIIPNNKYQ 575
 DB 1238 TLDOEKPELTLDNPKD---GKTKNEKTLVKGAV---SDN---LKD---YKVGKKA 1283
 QY 576 SLIGTQWHPEDLVYDIIRMEDEKKEVIRPTNMLTKTYTGLAGRTKDFHEI----- 627
 DB 1284 TVADGSYSARILLENGNEELKVIATDLAGKTKTKTYI-----DVNDRKVISGLIP 1335
 QY 628 -ELKNNKQELLISQTVTKDKTNLEPKGDKATF-----NLKHGESLTQ-----GLPEG 673
 DB 1336 GEDKNLK--AGESVKAIFASADLDTFTIRPLTNARASVONATELPLREISPGRYEG 1392
 QY 674 YSYLVKTEDESGYKVV 690
 DB 1393 YWTATSSIRKAKKAVEY 1409
 RESULT 4
 TOXB_CLODI STANDARD; PRT; 2366 AA.
 ID TOXB_CLODI
 AC P18177;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Toxin B.
 GN TOXB OR TCDB.
 OS Clostridium difficile.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OC NCBI_TaxID=1496;
 RN 11
 RN SEQUENCE FROM N.A.
 RC STRAIN-VPI 10463;
 RX MEDLINE=90326540; PubMed=2374729;
 RA Barroso L.A., Wang S.Z., Phelps C.J., Johnson J.L., Wilkins T.D.;
 RT "Nucleotide sequence of Clostridium difficile toxin B gene";
 RL Nucleic Acids Res. 18:4004-4004(1990).
 RN 12
 RN SEQUENCE FROM N.A.
 RC STRAIN-VPI 10463;
 RA von Eichel-Streiber C.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN 13
 RP SEQUENCE OF 1271-2366 FROM N.A.
 RC STRAIN-VPI 10463;


```

FT DOMAIN 2706 2811 CADHERIN 26.
FT DOMAIN 2812 2920 CADHERIN 27.
FT DOMAIN 2921 3024 CADHERIN 28.
FT DOMAIN 3025 3127 CADHERIN 29.
FT DOMAIN 3128 3232 CADHERIN 30.
FT DOMAIN 3233 3337 CADHERIN 31.
FT DOMAIN 3338 3442 CADHERIN 32.
FT DOMAIN 3443 3546 CADHERIN 33.
FT DOMAIN 3547 3649 CADHERIN 34.
FT DOMAIN 3650 3752 EGF-LIKE 1.
FT DOMAIN 3753 3855 LAMININ G-LIKE.
FT DOMAIN 3856 3958 EGF-LIKE 2.
FT DOMAIN 3959 4061 EGF-LIKE 3.
FT DOMAIN 4062 4164 EGF-LIKE 4.
FT DOMAIN 4165 4267 EGF-LIKE 5.
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 740 740 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 791 791 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 998 998 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1426 1426 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1551 1551 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1751 1751 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1867 1867 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1905 1905 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1943 1943 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1994 1994 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2328 2328 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2467 2467 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3326 3326 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3424 3424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3446 3446 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3615 3615 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3642 3642 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3718 3718 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4154 4154 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 4590 AA; 506273 MW; 04483CCD000E60A7 CRC64;

```

```

Query Match 3.58; Score 136.5; DB 1; Length 4590;
Best Local Similarity 19.88; Pred. No. 34;
Matches 176; Conservative 118; Mismatches 316; Indels 281; Gaps 46;

```

```

OY 47 AKTVEGLVESSTPAI---NPDSSSEYRWYGE-----SVYRGHPYKQFVADLRNL 98
DB 2573 AKDAGGKAVACTVAVILTDNDNAPOFRATYEIVNIGSSAAKGTSVYKASADAD----- 2626
OY 99 EGSRSYOVYCF---NIKRAFPGLSDSVYKWKYKHDGISTKFEEDYAMSPRITGDELNQ 153
DB 2627 EGSNADITVAIEADSESVKRENLKSLGVTITKESLIGLENEFTFE----- 2674
OY 154 KLRVAMVNGHQNANGIN-----EGLEP-----LNAIRVTOE 185
DB 2675 -VRV-ONGSPSKSVLVYVYKILPPEMQLPKFESEPYFTVSEDPVGETIDIRAEHS 2732
OY 186 AWVYYS---DNAPISNPDESFKRESSENLVST-----QLSLRQALQOLIDPN 231
DB 2733 GTVLYSLVKGKTPNSNDESFVIDROGRKLEKSLDHTTKWQFSLACIQD--DHE 2790
OY 232 LATMPKQVDDFOLISFESEDKGDKYNGK-----YONLLSGLV-----PKRPTPG 279
DB 2791 MVASV-----DVSIOVKDANDNSPVFESSPEAFIENLPGSGRVIQIRASDADSGTNG 2844
OY 280 DPRPMPNOPTSTVLIRKRYAGDYSKLLEGATL-QLGDNVNSQARVFSNDIGEREL 338
DB 2845 QVMSYSDOSQVEV-IESFALNMTGWI--TTLKLDELHKKDNDQIRKVAS-DHGEKIQL 2900
OY 339 SDGT---YTLLELNSPAGYSIAEPITKEV-----AGKYVYTI-----DGKQIE 379
DB 2901 SSTAIIVDVYITVDND-----SPRPTAETIKCTVSEDDPQGGVIALITLTDADSEIT- 2952
OY 380 NPNKEIVPYSEVAYNDPEESVLTTONYAKFYAKKNGSSOVVYCFNADLKSPDSED 439

```

```

DB 2953 --NRQVT--YFTGGDPLGQFAVEITIONEMKRY-----VKRPLDREK 2990
OY 440 -GKMTPEPDDFTTG-----EVKY-----THIAGRDLEKTYVKRPTDPTFLKHKK 484
DB 2991 RDNVLLTITAVDGFSSKAIVEKVLANDNSPCEKLTYSDTI-PEDVLP----- 3040
OY 485 VIEKGYREKGOAIEESGLTETQLNAATOLAIFYFTDSAE-----LDKDK 528
DB 3041 -----GKLIMQISATDADINSMAEITTYTLGSGAKFKRLNPDCELKTSRPLDDE 3091
OY 529 LKDYH-----GFG-----DMNDST--LAVAKLIVEADSNP-----POL 561
DB 3092 QAVHLLVLRATDGGRCFOASIVVLEEDVNDNAPEFSADPAITVFENTEPTGLLTRVOA 3151
OY 3152 TDLDLFFIPNNKYSGLT---QNHPELVLIIRNE---DK-----EVIPTNHLTL 608
DB 609 RKTVTGLA-----GDRTKDFHEIELKNNKOEILSOT-----VKTDTNLEF--- 650
OY 3211 RLATGTYIVSVLDINPNPVEFEKREYCATVSEDLVGTVELOYVYASRDLEANAETYS 3270
DB 651 -----KDGKATINLKHGSLTLQGL---PEGSYLVKETD-----SEGKRVVNSQEV 695
OY 3271 IISGNEHCKFSIDSKTGAVFILLENLDYESHSEYILVETATDGTGPSLSDVATVNVNTDI 3330
DB 696 ANAT-----VSKTGITSEDLFAENKKEPVVPGVDOKINGYIALIYIAG 740
OY 3331 NDNPNPVSQDYYTIVTYSIDAVAL--EQSVITVWADADBPNSHYSIIDG 3379

```

RESULT 7

```

SLAP_BACSH
ID SLAP_BACSH STANDARD; PRT: 1176 AA.
AC P38537;

```

```

DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Surface-layer protein 125 kDa protein precursor.
OS Bacillus sphaericus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1421;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2362;
RX MEDLINE=89327128; PubMed=2666389;
RA Bowditch R.D., Baumann P., Yousten A.A.;
RT "Cloning and sequencing of the gene encoding a 125-kilodalton
RT surface-layer protein from Bacillus sphaericus 2362 and of a related
RT cryptic gene."
RL J. Bacteriol. 171:4178-4188(1989).

```

```

CC -!- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC -!- OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.
CC -!- SUBCELLULAR LOCATION: Cell wall.
CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).

```

```

CC EMBL; M28361; AAA50256.1; -.
CC PIR; A33856; A33856.
CC InterPro; IPR001119; SLH.
CC Pfam; PF00395; SLH; 3.
CC PROSITE; PS01072; SLH_DOMAIN; 2.
CC Signal; Cell wall; S-layer; Repeat.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1176 SURFACE-LAYER 125 kDa PROTEIN.
FT DOMAIN 32 91 SLH 1.

```


FT DOMAIN 92 151 SLH 2.
 FT DOMAIN 152 210 SLH 3.
 SQ SEQUENCE 1176 AA; 125225 MW; 1A9A458EF43788C CRC64;

Query Match

Best Local Similarity 18.5%; Pred. No. 5.4;
 Matches 135; Conservative 101; Mismatches 256; Indels 238; Gaps 31;

166 NANGIMEGLEPL-----NAIRYTOEA-----WYYSNAPISNPDESFKR 205
 59 DANGNFPNPKTISRAREATIFTNALIELEAGDVNFKQVADANY-DATAIYENGIFEG 117
 206 ESSSNLVYSQSLMROALKOLIDPNLATMPKQVPDDFQSLFSESDKD----- 256
 118 VSATEPAPNKOLT-RSEPAKILVD-----AFEELEGEDLSEFADAST 158
 257 --KYNKGYNLSSGLVPTKPPPTGDPMPNPQPTSVLIRKAYAGDYSKLEGATLQ 314
 159 VKPMASYLEIAVANGYIKGSEANGKTNLPNAPTR---QDPAY-VFSRTLENDATP 213
 315 TGNVNSFOARFESSNDIGERIELSDGTTLTLETSAPAGSIAEPIFEKV-----EAGKV 369
 214 KVDKIEVDAKTLN-----VTLSDGTRETYLERALEPNKEVEYFKIKDVEYKAKVT 266
 370 YTIIDGKOIENPKKEIPEYPSVAYNDEEFVLTQNYA-----KFYAKNNGS 420
 267 YVTTATAVKSATNLKEVYVEFDGTVDKETEDANAYALKSGKITKYSVLAADNKTTAT 326
 421 SQVYVCFN---ADLKSPDESDGKTM---TPROFTGEVAYTHIAG-RDLFKYTVKRODT 473
 327 VTLTDLKNNKKAISISNVKAGDKETVAVNKVETAVDNKIPEVTEVKSIGTKAVKVTLS 386
 474 DPDTFLKIKKIVIEKGYREKGOAIEYSGLETOLEAATOLAIYFTDSALDKDKL---- 529
 387 EP-----VENLSTNTTLDGA--YFGNVMGAGNKTYLTPYSSALSVDGKHLVSG 438
 530 -KDYHGHGDMN-----DSTLAVAKLIVEYADDSNPOLTDIDFTIPNN 571
 439 AKDFAGFVSLNSTHEFEKVEVDKAPVTEATLFEVTLTFSED-----IDMDTVKASN 492
 572 -----NKYOSLI-GNQ-----WHPEDLVDI-----IRMDKR 597
 493 VYKSGDSKKEASEFERIADNKTKFYKFSSEKTLPTGKVDVYVEDIKDYSNKTAKDTKY 552
 598 EVIP-----VTNHLTKRTVTGLAGDRPKDFHFEIETLKNKOEILSQ 640
 553 TVPELIDQTRPEYKRYTALDEKTIKTYFESKTVDGESAIKYG--NYTKKDKD-----V 604
 641 VKTDKTNLEFKDGA-----TINKHGESLT-LOGLEPGYSYLVKETSDEG 685
 605 VSVDKVTYVDSKDSVYIIDLYSKVSGENTITIKNVKADATKLNMTMLDYGTFRSDKEG 664
 686 -----YKRYKN-----SOEAAKNTVSKTG- 704
 665 PDYEHVINDAKARKVYLFKDKKMDAASLADYSNYLVKINDTLQTLSEDAVATLSVNDAT 724
 705 ---ITSDETL 711
 725 VVITITFAETI 734

RESULT 8
 SLAP_BACST
 ID SLAP_BACST STANDARD; PRT; 1228 AA.
 AC P35825:
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE S-layer protein precursor (Surface layer protein).
 GN SBA.
 OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Geobacillus.
 NX NCBI_TaxId=1422;

FN [1]
 RC STRAIN=PV72;
 RX MEDLINE=94320770; PubMed=8045409;
 RA Kuen B., Sleytr U.B., Lubitz W.;
 RT "Sequence analysis of the sba gene encoding the 130-kDa
 surface-layer protein of Bacillus stearothermophilus strain PV72.";
 RL Gene 145:115-120(1994).
 CC -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
 CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 CC S-LAYER WITH HEXAGONAL SYMMETRY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

Query Match 3.4%; Score 135; DB 1; Length 1228;
 Best Local Similarity 19.3%; Pred. No. 5.8;
 Matches 128; Conservative 88; Mismatches 230; Indels 216; Gaps 33;

206 ESSSNLVSTQSLMROALKOLIDPNLATMPKQVPDDFQSLFSESDKDGNKNGYON 264
 551 EKESGVVASE-----LTYNADAKNVTLPK--ADLKENTTYQIKIKGLKSDGITE- 600
 265 LLSGLVPTKPPPTGDPMPNPQPTSVLIRKAYAGDYSKLEGATLQTL----- 315
 601 -----LGTVEKIEYERKTQDLAPYISVSKN--GD-----AGAKTTEAGFEYKVF 645
 316 GDNVNSFOARFESSNDIG-----ERIELSDGTTL----- 345
 646 SEMINFNATVVGSTIYGVAVVAKGANLSALTASDIIPASVEAVTGDDGYKAKVAA 705
 346 TELNSPAGYSI-----AEPIFEYKAGK-----YTTI-IDGKOIENPKKEIPEYPSVE 392
 706 NOLERNQYKLVFVGKATAPVADANANTLATNTYTTTTEGODVTAPT--VTKVFKGD 763
 393 AYNDPEEFSVLTQNYAKFEYAKNKGSSQVYVCFNADLKSPDESDGKTMTPDFTGE 452
 764 SLKDADAVTTL-----NVDAQOKFTIQFSEBELKTSSGSLVGKVT----- 804
 453 VKYTHIAGRDLEFYVYKPRDTPDLKIKKIVIEKGYREKGOAIEYSGLETOLEAATQ 512
 805 -----VEKLTNGMWDAGTGTVSAPKTDANGKYT 835
 513 LAIYFTDSAEIKD-KLK-----DYHGGKMDNTSLAVAKLIVEY-----A 553
 836 AAVYVTLTGLDNNKDKAKRLRYVDKSDTDGADAGAVNIEKDLILIRYSNMRHTVASAKA 895
 554 QDSNP-----POLDLDFEIPNNKQYSLISTGOMHPEDLVDIIMEDEKKEIYPTVHNL 606
 896 ADKDGOMASAPFTSTAID-----TYSLSL-VERNEEDLAEV--KEENIVKRAAQN 944
 607 TLKRTVTGLAGDRKDFHEIETLKNKOEILSQTV-----KT 643
 945 AVAGVTYALDGS-TNKFVF-----TPSQELKAGTVSVYIDGVRDKVGNITISKYTSFKT 998
 644 DKTN-----LEFDGKATINLKHGESLTQ-----GLPEGYSYLVKETSDESGYK--YKYN 691
 999 VSNAPTLSSISIADG--AAVNVRSKTTITLFSDSVNP-TITLTKRADGTSFTNYTLNVN 1055

Qy 692 SOEVANATVSKTGITSELTAFE--NNKEPVPTGVODK---INGYIAL--IVIGISLG 744
 Db 1056 NENNTYKIVFHKGVTLDEFTOYELAVSKDFOTGTDIDSKYFITIGSATDEVKALVGVG 1115
 Qy 745 IW 746
 Db 1116 SW 1117

RESULT 9
 SACS_YEAST STANDARD: PRT: 1301 AA.
 ID SACS_YEAST STANDARD: PRT: 1301 AA.
 AC P46674;
 DT 01-OCT-1995 (Rel. 32, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Leucine permease transcriptional regulator.
 GN SACS OR LEPI OR YDR159W OR YDR358.13.
 OS *Saccharomyces cerevisiae* (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; *Saccharomycotina*; *Saccharomycetes*;
 OC *Saccharomycetales*; *Saccharomycetaceae*; *Saccharomycetes*.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97027306; PubMed=8873450;
 RA Bauer A., Koelling R.;
 RT "Characterization of the SACS gene of *Saccharomyces cerevisiae*.";
 RL Yeast 12:965-975(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=S288C / AB972;
 RA Murphy L., Richards C., Harris D., Barrell B.G., Rajandream M.A.,
 RA Walsh S.V.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 189-1301 FROM N.A.
 RC STRAIN=DEY939;
 RA Stella C.A., Korch C., Ramos E.H., Mattoon J.R.;
 RT "Cloning and sequencing of LEPI, a gene associated with leucine
 transport.";
 RL Yeast 11:460-460(1995).
 CC -1- FUNCTION: POTENTIAL REGULATOR OF LEUCINE PERMEASE GENE(S)
 CC EXPRESSION. REQUIRED FOR NORMAL MITOSIS.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- SIMILARITY: BELONGS TO THE SACS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 247805; CAA87767.1; -;
 DR EMBL: 250046; CAA90379.1; -;
 DR EMBL: U35227; AAA79056.1; -;
 DR PIR: S51323; S51323.
 DR SGD: S0002586; SACS.
 DR GO: GO:0005643; C:nuclear pore; IDA.
 DR GO: GO:0005634; C:nucleus; IDA.
 DR GO: GO:0030029; P:actin filament-based process; IGI.
 DR GO: GO:0007067; P:mitosis; IMP.
 DR GO: GO:0006611; P:protein-nucleus export; IGI.
 DR InterPro: IPR005062; SACS_GANP.
 DR Pfam: PF03399; SACS_GANP; 1.
 DR Transcription regulation; Nuclear protein.
 KW SEQUENCE 1301 AA; 149568 MW; 0679DB1673DDACB CRC64;

Query Match 3.4%; Score 135; DB 1; Length 1301;
 Best Local Similarity 19.1%; Pred. No. 6.3;
 Matches 139; Conservative 105; Mismatches 307; Indels 176; Gaps 31;

Qy 115 PFLGSSSVKMKYKHKHDISTKPEYAMSPRTGD---ELNOKLRAVMYNGHP---ONA 167
 Db 611 PPSQESQQLPOLQSOSHLST---NPLTFPOVHGLDSKOKQOIKTVTDGSSPPVFDOSA 667
 Qy 168 NGIMEGLEPLNAIRVTOEAVVYSDNADISPDSFKRESNSLVTSQLSIMQ-ALKQ 226
 Db 668 QNSTVEASKAMHISTTSNGAY---DEKLSQEEKRRKEEORIEBEKQTLKKQENADQ 724
 Qy 227 LIDRLATKMPKQVPDDPOLSLFESQKGDKNKCYQLLSGGLVPTKPPPGDPMPPN 286
 Db 725 VITEQIANDLVKEVNVSSVIVIKREFSEAVNRQDFDTWTRLY-----DAFLER 776
 Qy 287 -----OPQTSVLIRK-----VAIGDSKLE-----GATLOTGDN 318
 Db 777 LYLIMDSRALKRNSTLKKKFFEKWASQAKKRNRLLEKKEKEITLVSHQLGPGFK 836
 Qy 319 VNSFOARVSSNDIGERIELSDGYTTLTSLNSPAGYSIAEPITKEAGKYVTTIDKQI 378
 Db 837 KSTCLFRPYKGVNNSFMLSDDKNL--IFSPVDEFNKPATHLTKISKLRPLEMQSI 894
 Qy 379 --ENPNKEI---VEPVSEAYND-----FEERSVLTQNYAKFYAKKNKSSQ 422
 Db 895 YYDNLTKFPNSNLTLPANLFTYADWTSLSNRWILSKENLOTADSKKF---SNNIIS 950
 Qy 423 VVYCFNADLKSPPOSED-----GKTMPTDFTTGEVYK-----THI 458
 Db 951 RIICID-DEYSPSPSDQLILFNTGYT-NFDIDLEKLKDDGEELIKLTGISTLNIN 1008
 Qy 459 AGRDLFKYTVKRPDTPDPTFLKIKV--IEKGYREKQALYSGLT-----EQQLR 508
 Db 1009 CFSLLIYWEASAEMLTSESTIKHLKLNRSKNYSVIERIDMLMFEESPDKLEKLS 1068
 Qy 509 AATQLATIYFTDSAEIDKDKLDYHGFDMDNSTLAVAKLIVEAODSNPQL-DLD-- 565
 Db 1069 EISHSYVKLTGERGYDK-----TLROKRSIAGIHSRSTQLQTKRKIDQK 1113
 Qy 566 --FFIPNNKYSQSLIG---TQWHPEDLVDIRMDEKKEVLPV---THN-----LTIR 609
 Db 1114 MKKMEKRNKNYQOQIGERNYVAHLESHID-ASPRSKRKRPILISTSHSSQFPTPLASR 1172
 Qy 610 KTVTGLADRTKDFEIELEKNNKQELLSTQVTKDTKLTLEKRD-----GKATINLKHG 662
 Db 1173 LNTGSSSTSPPLPSHLAKFKRNSRVTSLHTVLPVSTPSHNNIPASFSGNNTTDI--- 1229
 Qy 663 ESLTLGCPGEGSYLVKRTDSEGYKVKVNSDEVANATYSKIGTISDELTAEENKEPVVP 722
 Db 1230 -----QSQQLIENCKSTSVYLVNNYSERILGNOEICQPTI-----NPVTP 1268
 Qy 723 T--GVQD 727
 Db 1269 VLDGADQ 1275

RESULT 10
 SLPH_BRECH STANDARD: PRT: 1116 AA.
 ID SLPH_BRECH STANDARD: PRT: 1116 AA.
 AC P38538;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Surface layer protein precursor (Hexagonal wall protein) (HWP).
 OS *Brevibacillus choshuensis*.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Brevibacillus*.
 OX NCBI_TaxID=54911;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 54-68.
 RC STRAIN=HPD31;
 RX MEDLINE=90170842; PubMed=2307650;
 RA Ebisu S., Tsuboi A., Takagi H., Naruse Y., Yamagata H., Tsukagoshi N.,
 RA Uda S.;
 RT "Conserved structures of cell wall protein genes among protein-
 producing *Bacillus brevis* strains.";
 RL J. Bacteriol. 172:1312-1320(1990).

```

CC - SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC HEXAGONAL S-LAYER.
CC - SIMILARITY: Contains 3 s-layer homology (SLH) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its contents is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D90050; BAA14103.1; -
CC InterPro: IPR001119; SLH.
CC Pfam: PF00395; SLH; 2.
CC PROSITE: PS01072; SLH DOMAIN; 2.
CC Signal: Cell wall; S-layer; Repeat.
CC SIGNAL 1 33
CC CHAIN 54 1116 SURFACE LAYER PROTEIN.
CC DOMAIN 57 120 SLH 1.
CC DOMAIN 121 171 SLH 2.
CC DOMAIN 172 231 SLH 3.
CC SEQUENCE 1116 AA; 123397 MW; 86D583D7AC72546F CRC64;

Query Match 3.4%; Score 132.5; DB 1; Length 1116;
Best Local Similarity 18.9%; Pred. No. 7;
Matches 162; Conservative 112; Mismatches 266; Indels 315; Gaps 45;

QY 53 LVESSTPMAINPDSSEY-----RWGYESTYVG-----HPY 84
DB 143 IVKGFPPKSFPPQNVTAEAVTMIVRAGLGPSPVGNPNMISKSELINAKGINNP 202
QY 85 YKOFRA-----HDLRVNLEGSRYQVCFNLKAFPLGSD-----SSYKKWKKH 130
DB 203 MGOFAATIRKMDNALRVKLMEOIE-----GTRLRVNTDETILTKLK-- 247
QY 131 DGISTKEFDYAMSPRTGDEL-----NOKLRVYNGHPQONANG-----IMEG 173
DB 248 --VTYVDMQMAHEKGNNSDELPLYTVNPAIGLSLKANEVTLNGKADLGSNTTYKVAEG 305
QY 174 LEPINLRVYQEAHVWYSDNAPISNDESEFKRESENLVSTQSLMRQAKLIDPNLA 233
DB 306 INP-NAPDQKVQVWTKIDRENT---VMEGSEDEVDVMDVSAVLYKRGKFTD-DIV 359
QY 234 TPKMPQVDPDFLOFISEDEKGDKNYKGYONLLSGLVPTKPTPDGPPMPRQPTTSV 293
DB 360 KDLKSDLDVKIKEMGSE----- 378
QY 294 LIRKTAIGDYSKLBGATIQLT--GDVNSFOARFSSND-IGERLELSDGYTL----- 345
DB 379 --KSYRLTDEDTKI---TYNFTRENDPVDAI-SKIYKNDTDFGVKVVLDNNEVAYLHII 431
QY 346 ---TELNSPAGYSIAPIITFVEA-----GKYVTI-IDGKOIEN 380
DB 432 DQGITDKSKYKGYKSVISKIDADKKTKTINDNSKFSLEDDODEBKDPLVFDGPAKY 491
QY 381 PKKEIYPSV-EAYNDFEEFVLTQNYAKFYAKNKGSSQVYCFNADLKSPDSED 439
DB 492 GLKESDYVSYVYADGDKLYLFANRNAE-----GKYEKVSRN---KTDRLVY 540
QY 440 GKRT--MTPDFTTGKYYTHIAGRLFKTT---VKPND-----TPPDFLFLKIK-- 483
DB 541 GKRTYKVPD-----ASYSENANKDKKYNNSDLISNDGEEVKLLDPPSGVRHIEIK 595
QY 484 -----KVIKGG--YREKGALEYSLTETQTLAATQALAIYFTDSAEIDKDKLKOY 532
DB 596 DAIDRRKPLAITIKGFTYNSKQTDYFWTQ---KGRTOI-----VSLDQKIYDR 644
QY 533 HG--FGDMNDSTLAVKILVEYQA-----DSNPQTLIDLDF-----FIPN 570
DB 645 YGVNVDKSKNDKQAFKDLVQLQPKYVAKEDATDANQVLELVNDSNGEVDKQYVLD 704
QY 571 NNKYOSLIGTQW-----HPEDLVDIRMEDKEVIVPTIHNLTKLTKVTGLAGDRTKDHF 625

```

```

DB 705 KLYSE--KSTWMDKLADEDDVDVGYEDVDKTAVFMTGDLT---PATG-----TK----- 750
QY 626 EELKN-----NKDEL-LSQTVKNDKTLF----- 649
DB 751 RGLKNAKTAKEKVDKAKSDLKWYVSDKEKGEVQAIYVVDGSLGQHOPGWKQYGT 810
QY 650 FKDKKATINLKHGESLT-----LOGIPEGYSYLKKEETDESEGYKV--KVNQOEYANAFVK 702
DB 811 SKQDITITVTKQSDSTYKEKYKLDG-----DADDLKVDQDIRGVDISTFLNS 858
QY 703 TG-TTSDETLAEFNN 716
DB 859 DGEVIVDVEYEVN 873

RESULT 11
CADN_DROME
ID CADN_DROME STANDARD; PRT; 3097 AA.
AC O15943; Q9VJB7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neural-cadherin precursor (Cadherin-N protein) (DN-cadherin).
GN CADN OR CG7100.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Pnygroidae; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM D).
RC TRISSE-Embryo, and Head;
RX MEDLINE=97388431; PubMed=9247265;
RA Iwai Y., Usui T., Hirano S., Steward R., Takeichi M., Uemura T.;
RT "Axon patterning requires DN-cadherin, a novel neuronal adhesion
RL receptor, in the Drosophila embryonic CNS. ";
RN Neuron 19:77-89(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.D.G.,
RA Abirl J.F., Agdayani A., An H.-J., Andrews-Plankkoch C., Baldwin D.,
RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burks C.C., Busan D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazozo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

```

RA Wang Z.Y., Massarom D.A., Weinstein G.M., Weissensbach J.J.,
RA Williams S.M., Woodage T., Wooley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*.",
RL Science 287:2185-2195(2000).
RN [3]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RC STRAIN-Berkeley.
RX MEDLINE=22426069; PubMed=12537572;
RA Mista S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu U., Bertman B.P.,
RA Bettencourt B.R., Gelniker S.E., de Grey A.D.N.-J., Dysdale R.A.,
RA Harris N.L., Richter J., Ruso S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.,
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RT systematic review".
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [4]
RP INTERACTION WITH ARM.
RX MEDLINE=98298928; PubMed=9635189;
RA Loureiro J., Pelzer M.:
RT "Roles of Armadillo, a *Drosophila* catenin, during central nervous
RT system development".
RL Curr. Biol. 8:623-632(1998).
CC -I- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS: CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. MAY ASSOCIATE WITH ARM NEURAL
CC ISOFORM AND PARTICIPATE IN THE TRANSMISSION OF DEVELOPMENTAL
CC INFORMATION.
CC -I- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -I- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=8;
CC Comment=Experimental confirmation may be lacking for some
CC isoforms;
CC Name=D;
CC IsoId=O15943-1; Sequence=Displayed;
CC Name=A;
CC IsoId=O15943-2; Sequence=VSP_000667, VSP_000668;
CC Name=B;
CC IsoId=O15943-3; Sequence=VSP_000668;
CC Name=C;
CC IsoId=O15943-4; Sequence=VSP_000667, VSP_000668, VSP_000669;
CC Name=E;
CC IsoId=O15943-5; Sequence=VSP_000667;
CC Name=F;
CC IsoId=O15943-6; Sequence=VSP_000669;
CC Name=G;
CC IsoId=O15943-7; Sequence=VSP_000667, VSP_000669;
CC Name=H;
CC IsoId=O15943-8; Sequence=VSP_000668, VSP_000669;
CC -I- TISSUE SPECIFICITY: IN THE EMBRYO, THE PROTEIN FIRST APPEARS IN
CC THE MESOEM AT STAGE 9 AND IS PRESENT IN THE MYOBLASTS AND MUSCLE
CC FIBERS BY STAGE 12 AND STAGE 14, RESPECTIVELY. AT STAGE 12 THE
CC PROTEIN IS ALSO LOCATED IN THE AXONS OF THE ENTIRE CNS, BUT NOT IN
CC THE GLIAL CELLS. IN THIRD INSTAR LARVAE PROTEIN IS EXPRESSED IN
CC THE CNS NEURONELS, PHOTORECEPTOR AXONS AND PRECURSORS OF ADULT
CC MUSCLES.
CC -I- SIMILARITY: Contains 16 cadherin domains.
CC -I- SIMILARITY: Contains 3 EGF-like domains.
CC -I- SIMILARITY: Contains 2 laminin G-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
CC or send an email to license@isb-sib.ch).

CC	Protein	Accession	Gene	Function
DR	EMBL	AB002397	BAA22131.1	-
DR	EMBL	AE003656	AAF53635.1	-
DR	EMBL	AE003656	AAN10992.1	-
DR	EMBL	AE003656	AAN10993.1	-
DR	EMBL	AE003656	AAN10994.1	-
DR	EMBL	AE003656	AAN10995.1	-
DR	EMBL	AE003656	AAN10996.1	-
DR	EMBL	AE003656	AAN10997.1	-
DR	PIR	T00021	T00021	-
DR	HSSP	P00740	1EDM	-
DR	Flybase	FBgn0015609	Cadn	Calcium-binding protein; cell adhesion molecule ac. . ; IPI
DR	GO	GO:0005911	C:intercellular junction; IDA	-
DR	GO	GO:0008014	F:calcium-dependent cell adhesion molecule ac. . ; IPI	-
DR	GO	GO:0007412	P:axon target recognition; IMP	-
DR	GO	GO:0007156	P:homophilic cell adhesion; IDA	-
DR	InterPro	IPR002126	Cadherin	-
DR	InterPro	IPR000233	Cadherin_C_term	-
DR	InterPro	IPR000742	EGF_2	-
DR	InterPro	IPR01881	EGF_Ca	-
DR	InterPro	IPR006209	EGF_Like	-
DR	InterPro	IPR01791	Laminin_G	-
DR	Pfam	PF00028	cadherin; 14	-
DR	Pfam	PF01049	Cadherin_C_term; 1	-
DR	Pfam	PF00008	EGF; 3	-
DR	Pfam	PF00054	Laminin_G; 2	-
DR	PRINTS	PRO0205	CADHERIN	-
DR	SMART	SM0112	CA; 16	-
DR	SMART	SM0179	EGF_CA; 1	-
DR	SMART	SM00282	lamc; 2	-
DR	PROSITE	PS00232	CADHERIN_1; 9	-
DR	PROSITE	PS50268	CADHERIN_2; 16	-
DR	PROSITE	PS00022	EGF_1; 3	-
DR	PROSITE	PS01186	EGF_2; 3	-
DR	PROSITE	PS50025	LAM_G_DOMAIN; 2	-
KW	Cell adhesion	Glycoprotein; Transmembrane; Calcium-binding; Repeat;		
KW	Signal	EGF-like domain; Alternating splicing		
FT	SIGNAL	1	36	POTENTIAL
FT	PROPEP	37	?	
FT	CHAIN	?	3097	NEURAL-CADHERIN
FT	DOMAIN	?	1454	EXTRACELLULAR (POTENTIAL)
FT	TRANSMEM	1455	1475	POTENTIAL
FT	DOMAIN	1476	3097	CYTOPLASMIC (POTENTIAL)
FT	DOMAIN	181	305	CADHERIN_1
FT	DOMAIN	430	543	CADHERIN_2
FT	DOMAIN	554	651	CADHERIN_3
FT	DOMAIN	660	756	CADHERIN_4
FT	DOMAIN	766	858	CADHERIN_5
FT	DOMAIN	867	968	CADHERIN_6
FT	DOMAIN	978	1078	CADHERIN_7
FT	DOMAIN	1087	1183	CADHERIN_8
FT	DOMAIN	1193	1299	CADHERIN_9
FT	DOMAIN	1307	1414	CADHERIN_10
FT	DOMAIN	1423	1514	CADHERIN_11
FT	DOMAIN	1523	1630	CADHERIN_12
FT	DOMAIN	1639	1742	CADHERIN_13
FT	DOMAIN	1749	1861	CADHERIN_14
FT	DOMAIN	1870	1966	CADHERIN_15
FT	DOMAIN	1974	2085	CADHERIN_16
FT	DOMAIN	2346	2377	EGF-LIKE 1
FT	DOMAIN	2379	2585	LAMININ G-LIKE 1
FT	DOMAIN	2592	2627	EGF-LIKE 2
FT	DOMAIN	2631	2822	LAMININ 3
FT	DOMAIN	2859	2902	EGF-LIKE 3
FT	DISULFID	2346	2357	POTENTIAL
FT	DISULFID	2351	2366	POTENTIAL
FT	DISULFID	2368	2377	POTENTIAL
FT	DISULFID	2592	2607	POTENTIAL
FT	DISULFID	2601	2616	POTENTIAL
FT	DISULFID	2618	2627	POTENTIAL
FT	DISULFID	2869	2880	POTENTIAL

Query Match 3.4%; Score 132.5; DB 1; Length 3097;
 Best Local Similarity 19.8%; Pred. No. 32;
 Matches 114; Conservative 89; Mismatches 192; Indels 181; Gaps 29;

QY 287 QPOTSVLIRKKA-IGDYSLGALQLT--GDVNSFQARFSSNDICERLESDGT 343
 DB 719 QPHVYSILVATEDSDSGFSTVD-LTRIVDVNDNMFELPYQANVEDIDPLGTSIL 777
 QY 344 TLELNSPAG-----YSIAEPITFEKGVYTIIDGKQIENPKRIVEPSYEAVND 397
 DB 778 RVAMSDSGSNAEIELVSD--HFAVDSNGI--YANKQDLDNANNAYEFITAKDKG 834
 QY 398 E-EFSLTQNTAKFYAKNKGSS--SQVYCFENADLKSP-----DSED 439
 DB 835 EPPKSGVAVR-----VYTKNKNDDEPKFSQYTVPVNDNAGPNTLVTVASDKGD 889
 QY 440 -----GKTMTPDET-----TGEVKYTHA-----GRDLFKTYV----- 468
 DB 890 VREGVGGTSSQFVIEDITGVIRLHNKAISLDKRELYNVTAMDSCCVNGDQTIHT 949
 QY 469 -----KPRDTPDTFLKHK-----KVI-----EKGYREKQAIEX 499
 DB 950 STAVVVVETDVNDKRVFECSTYPRKVEGAPNGSPVILKVAATDEKGV--NGQ-VKY 1006
 QY 500 SGLTEQLRAATQALAIYFTDSALDKLKDYG-----FGDNDSTL-AVAK 547
 DB 1007 S-IVQGPNOGKFTV--DETEGEVSTNKVFDEGDKGFTSVYKATDGDPSLEGVCS 1063
 QY 548 ILVEYADSNRPOLTDLDFIPNNKYSQSLGTOMHEDVDIIRBDEKKEVLPVHNL 607
 DB 1064 FVEIETDVNDNPLFPRQKV--ENVKODASIGT-----NLRVS----- 1101
 QY 608 LRKTVGLADGRKDFEFIELEKNNKQELLSTQVTKDNTLEKDGKATINLKHESLTL 667
 DB 1102 -----ASDEAD-----NNGAIYSLTAPENPLEFEIQA----- 1133
 QY 668 QGLPEGSYLVKETDESEGVYKVNSEVANATVSKTGTSDETLAENKPEVPTGVD 727
 DB 1134 -----ESGMIVLKKPLDETYKLEAMADKGYPLSRFVEVDIVVDRANN-----PRYWDH 1185
 QY 728 KINGYLAL-----IYAGISLIGWGHITRIR 754
 DB 1186 TVYGPITYKENMPGVGKVASIKASSGLEGNPTVFYR 1221

RESULT 12
 LHS1_YEAST STANDARD; PRT; 881 AA.
 AC P36016;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Heat shock protein 70 homolog LHS1 precursor.
 GN LHS1 OR YKJ073W OR YKJ355.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Pohl T.M., Pohl F.M.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA STRAIN=5288C;
 RX MEDLINE=94378724; PubMed=8091863;
 RA Rasmussen S.W.;
 RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
 RT NUP100 gene, an open reading frame (ORF) possibly representing a
 RT nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
 RT addition to seven ORFs with weak or no significant similarity to
 RT known proteins";
 RL Yeast 10:569-574(1994).

RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=96256281; PubMed=8654361;
 RA Craven R.A., Egerton M., Stirling C.J.;
 RT "A novel Hsp70 of the yeast *Er* lumen is required for the efficient
 RT translocation of a number of protein precursors";
 RL EMO J. 15:2640-2650(1996).
 CC -1- FUNCTION: PROBABLY ACTS AS A CHAPERONE INVOLVED IN BOTH
 CC POLYPEPTIDE TRANSLOCATION AND SUBSEQUENT FOLDING.
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
 CC -1- PTM: N-GLYCOSYLATED.
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X75780; CA53401.1; -;
 DR EMBL: Z28073; CA81910.1; -;
 DR PIR: S37895; S37895.
 DR SGD: S0001556; LHS1.
 DR GO: GO:0005783; C: endoplasmic reticulum; IDA.
 DR GO: GO:0003754; F: chaperone activity; IMP.
 DR GO: GO:0015031; P: protein transport; ICI.
 DR GO: GO:0006986; P: response to unfolded protein; IGI.
 DR InterPro: IPR000886; ER_target.
 DR InterPro: IPR01023; Hsp70.
 DR Pfam: PF00012; Hsp70; 1.
 DR PRINTS: PR00301; HEATSHOCK70.
 DR ProDom: PD000089; Hsp70; 1.
 DR PROSITE: PS00014; ER_TARGET; 1.
 DR PROSITE: PS00329; HSP70_2; 1.
 DR PROSITE: PS01036; HSP70_3; 1.
 DR PROSITE: PS00297; HSP70_1; FALSE NEG.
 KW Chaperone; Endoplasmic reticulum; Glycoprotein; signal; ATP-binding.
 FT STGNL 1 20
 FT CHAIN 21 881
 FT CARBOHYD 128 128
 FT CARBOHYD 458 458
 FT CARBOHYD 474 474
 FT CARBOHYD 481 481
 FT CARBOHYD 489 489
 FT CARBOHYD 527 527
 FT CARBOHYD 844 844
 FT SITE 878 881
 SQ SHOUENCE 881 AA; 99571 MW; ACED092CA3A34785 CRC64;

Query Match 3.3%; Score 130.5; DB 1; Length 881;
 Best Local Similarity 18.3%; Pred. No. 6.4; 270; Indels 199; Gaps 32;
 Matches 132; Conservative 119; Mismatches 199;

QY 52 GLVESSTPNAINDSSS-----EYRWYGESYVRG-----HPYKOF 88
 DB 241 GSIKAMFSLQPEDTTPVTEFEGYVPHGAKFTMDISLKNFLEHNPALR-- 298
 QY 89 RVAHDLRVNIEGSRSYQVCFNLKKAAPPLGSDSSVKYKKWKHKDGTCKEDYAMSRTIG 148
 DB 299 --TDELHANPKALAKINAQAEKAKLILANSSEASI-----NIESLINDIDFRSTIR 348
 QY 149 DELNOKLRVAMVNGHQNANGIMELEPNAIRVTOEAVWYSDNAPISNPDSFKREE 208
 DB 349 QEFEEFT-----ADSLDIYKPLIND-AVTKQ-PGCGTNP-----E 383
 QY 209 SNLV-----STQSLNRQALQQLIDPNLTKMKRQVDDFOLSISEEDKDKYNGYON 264
 DB 384 INVILLAGSSRRPIYQDLIKYSEE-----KVIARNAP-ESAVNGVWKGITLSNFK- 438
 QY 265 LLSGLVPTKRPPTGPPMP-----PNQPTTSVLIRKVAIGDYSLLEGATLQJGN 318

```

Db 439 -----TKPLNVDSRVNTYSEFKLSNESELYDFTRGSAVPNKTSILTNT-----DS 485
Oy 319 V-NSFOARVSSNDIGERIELSDGTYYITTELNSPAGYSIAPPIPKVAGVYTTID--- 374
Db 486 IPNNFTIDLENGKLFETTYVNSAINKNSYSSDKSCSSGVAANTFEDLSDBLFSIQEYNC 545
Oy 375 -----GKQIENPNK-----EIVEPYSEVAYNFEESVLTQONVAKFYAK 415
Db 546 ICOSENDIGNSKQIKNKGSRLATFSEDEVEIKRLSPSESRLEHKKLKDQKDEFCQOE 605
Oy 416 NKNGSSOVVYCFNADLSPDSEDCGKTMPDFTGVEKYTHIAGRDLEKTYVAPRDTDP 475
Db 606 NLNLESNLXDARNLMDDEVMQNGKQVEELS--EMVKYLLDMLLEDAF-----DTPD 658
Oy 476 DTFPKHT-----KKVIE-----KGYRKG-----QAIYSGLTETQLAA 510
Db 659 EDVYSRIREIGILKKLTELVDYSAKEPINSOQFGMEEGHKLLOAITHRNTEEFISQ 718
Oy 511 TOLAIYFTDSAEIDKDKLDYHGFDMNDSTLAVAKILVEYAO--DSNPQLDPLDFEI 568
Db 719 FETFEADTIDNVREFFKKIK-----QPAYSKALSTWEEITLTSFKNSISIEKFL 768
Oy 569 PNNKYSOLIGTQMHPELDVDI-----IMEDKKEVIVTHNLTKTGTGLAGDRT 620
Db 769 AKN-----LFGEDLR-EHLFEIKLOFDWYRTKLEKRLIK-----SGDES 808
Oy 621 -----KDFHF-EIILKNNKOLLSTQV-----KIDKTNLEKCKOKAT-NKHS 664
Db 809 RLNEIKLHLNRFLOKREKRLKLEOKSRNNNETESTVINSADKTTIYNDKTES 868

```

RESULT 13

```

DPOL_BORBU STANDARD; PRT; 908 AA.
ID DPOL_BORBU
AC 051498;

```

```

DT 30-MAY-2000 (Rel. 39, Created)

```

```

DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

```

```

DE DNA polymerase I (EC 2.7.7.7) (POL I).
GN POLA OR BB0548.

```

```

OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
NX NCBI_TaxId=139;

```

```

RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;

```

```

RX MEDLINE=98065943; PubMed=9403685;

```

```

RA Fraser C.M., Caajens S., Huang W.M., Sutton G.G., Clayton R.A.,
Lalibiera R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,

```

```

RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
Peterson J., Katiavage A.R., Quackenbush J., Salzberg S., Hanson M.,

```

```

RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
Uterback T., Matthey L., McDonald L., Artlisch P., Bowman C.,

```

```

RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
Smith H.O., Venter J.C.;

```

```

RT "Genomic sequence of a Lyme disease spirochete, Borrelia
burgdorferi."

```

```

RL Nature 390:580-586(1997).

```

```

CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY (BY
SIMILARITY).

```

```

CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
+ (DNA)(n).

```

```

CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-A FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb-sib.ch).

```

```

DR EMBL: AE001156; AAC66909.1; -
DR PIR: C70168; C70168.
DR HSSP: P19821; 1KTQ.
DR TIGR: BB0548; -.
DR InterPro: IPR002562; 3_5_exonuclease.
DR InterPro: IPR002421; 5_3_exonuclease.
DR InterPro: IPR001098; DNA_POL.
DR InterPro: IPR002298; DNA_POL.
DR InterPro: IPR000513; Exo_N_I.
DR InterPro: IPR003584; HNH_2.
DR Pfam: PF01612; 3_5_exonuclease; 1.
DR Pfam: PF02739; 5_3_exonuclease; 1.
DR Pfam: PF01367; 5_3_exonuclease; 1.
DR Pfam: PF00476; DNA_POL_A; 1.
DR PRINTS: PR00668; DNAPOL1.
DR SMART: SM00474; 3EXOC; 1.
DR SMART: SM00475; 53EXOC; 1.
DR SMART: SM00279; HNH2; 1.
DR SMART: SM00482; POLAC; 1.
DR TIGRfam: TIGR00593; polA; 1.
DR PROSITE: PS00447; DNA_POLYMERASE_A; 1.
KW Translase; DNA-directed DNA polymerase; DNA replication; DNA repair;
KW Hydrolase; Exonuclease; DNA-binding; Complete proteome.
FT DOMAIN 1 318
FT DOMAIN 319 531
FT DOMAIN 532 908
SQ SEQUENCE 908 AA; 105503 MW; B58512AE80C8BD94 CRC64;

```

```

Query Match 3.3%; Score 130.5; DB 1; Length 908;
Best Local Similarity 19.3%; Pred. No. 6.6;
Matches 166; Conservative 103; Mismatches 264; Indels 327; Gaps 43;

```

```

Oy 13 LNTQVLSKNSKRFVTYLVGFALFALVTSWGAKTYVGLVESSTPAINPDSSEFRW 72
Db 26 LNTQ---GENVNAF-----IGFF-----KTLFPIKEKNPEHLITTFDESVPT 65
Oy 73 YGESYVAGHPYKQFR-----VANDLRVNL--- 98
Db 66 FRKQKTY---PSYKATRLPPDDLIPQIGWIKKGLAKKPIFEEMEGEADDLASFARK 121
Oy 99 EGSRSYQVYCFN-----LKKAPPLGSSSVK---KWKKHDCI-STKPEDYAMS 143
Db 122 AAKNNYLYIISPOKDLQTMSEYVKILKIENSNFEMDNVEYTKKFCVNSQIOLDYL- 179
Oy 144 PRITGDELNOKLRVAVMNGHPOMAN-----GIMGLEPLN-----AIRVQEVW 188
Db 180 -AIVGDR-SDNIPGKIGAGGAANLREFYTLDSIYNLEITNKKHRELLIKERENAF 237
Oy 189 YV-----SDNAPISMPD-----ESFKRESESNVSTQSLSMRQALKOLIDPLNATKMKQ 239
Db 238 STELVSELENKIPETIENFALKNESEELISLEPKHSATALLKTYKKDIL-----KQKE 291
Oy 240 VPDDFOLSFSESEDKGRYKNKYQNLGSLGVTPKPTPGDPMPBPNOPTTVLIRKYA 299
Db 292 NAD--QKSLFKQE-----PTNLSL----- 308
Oy 300 IGVSKLLEGATLQLTGNNVSPQARVSSNDIGRIELSDGTTL-----TEINS 350
Db 309 -----DDIMTIDTENVKYRSITFKIELDLIDLELKRKAKYISIDTERSS 351
Oy 351 PAGYS---IAEPIPK-----VEA-GKYVYTIIDGQIENPNKEIPEYSEVAYND-F 397
Db 352 LDYTAKLIGISISKEKEGYIPIEAKGKIT-----IEKNYTIQFNMLF 397
Oy 398 EEFVLTQYNTAKFYAAKNGSSQVYVCFNADLSPDSEDCGKTMTPDF---TTGEVK 454
Db 398 ESNPRLIQGN-KPDYKILKNG-----FN---PIPPYD---TMAAVLIDNSVVS 443
Oy 455 YTHIAGRLPFTYVAKPRTDPTFLKHKYIEKYEKRGGAIEVSGLTQLARATOLA 514
Db 444 LDPLAEKYLHMKNIKIED-----VIQKNDNFANISLEWATSYSSBDADITPRL 491
Oy 515 IYFTDSAEIDK-DKLKDYHGFDMNDSTLAVAKILVEYAO----- 555

```

DB 492 FNIFTKKEDKDLK-----MHEIEMFNKVIIEEMENCIYLDKEYLKEYGRELGR 543
 OY 556 -----SNPOLND-----LDFIPNNKKQSLIGTQWHEEDLVDDI 591
 DB 544 ELEAIENEIKSIGIDFNPNSPKOMHEILFEKLNKLPEKMK-----KSDTDIK 592
 OY 592 RMEDKKEVLPVHTNLTLRKTVTGLAGDRTKDFHFEIEELKNNKOELLSTQVTKDTNLEFK 651
 DB 593 VLESLEQHESEIENLTKYHQIAKLKSTYTDNL--LELIN-----YKTNRLHLSFI 640
 OY 652 DQKATINLKHGSELTLOQPEGSIYKVEDSGYIV-KVNSEOVANATYK-----702
 DB 641 QRTATGRTSTINPNLQNTIP-----IK--DEKGRRIKRAFKPENGNIFISADYSQIEEA 692
 OY 703 --TGTSDETL--AFENKKE 718
 DB 693 ILAHLSDDEVILKAFENKMD 712

RESULT 14
 RPGR_HUMAN
 ID AC Q92834; UNMAN STANFORD; PRT: 1020 AA.
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE X-linked retinitis pigmentosa GTPase regulator.
 GN RPGR OR RP3 OR XLRP3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4), AND VARIANTS RP3 GLN-98;
 RP VAL-215; ARG-250 AND 296-THR--ILE-300 DEL.
 RX MEDLINE=96241570; PubMed=8673101;
 RA Meindl A., Dry K.L., Herrmann K., Manson F.D., Ciccodicola A.,
 RA Edgar A.J., Carvalho M.R.S., Achatz H., Hellebrand H., Lennon A.A.,
 RA Migliaccio C., Porter K., Zrenner E., Bird A.C., Jay M., Lorenz B.,
 RA Wlter B., D'Urso M., Mellinger T., Wright A.;
 RA "A gene (RPGR) with homology to the RGC1 guanine nucleotide exchange
 RT factor is mutated in X-linked retinitis pigmentosa (RP3).";
 RL Nat. Genet. 13:35-42(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANTS RP3 CYS-130; SER-235 AND
 RP SER-275.
 RC TISSUE=Retina;
 RX MEDLINE=96414315; PubMed=8817343;
 RA Roepman R., van Duijnhoven G., Rosenberg T., Pinckers A.J.L.G.,
 RA Bleeker-Wagemakers L.M., Bergen A.A.B., Post J., Beck A.,
 RA Reimhardt R., Ropers H.-H., Cremers F., Berger W.;
 RA "Positional cloning of the gene for X-linked retinitis pigmentosa 3:
 RT homology with the guanine-nucleotide-exchange factor RGC1.";
 RL Hum. Mol. Genet. 5:1035-1041(1996).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
 RC TISSUE=Retina;
 RX MEDLINE=99330567; PubMed=10401007;
 RA Kirschner R., Rosenberg T., Schultz-Hendbrok R., Lenzner S., Feil S.,
 RA Roepman R., Cremers F.P.M., Ropers H.-H., Berger W.;
 RA "RPGR transcription studies in mouse and human tissues reveal a
 RT retina-specific isoform that is disrupted in a patient with X-linked
 RL retinitis pigmentosa.";
 RL Hum. Mol. Genet. 8:1571-1578(1999).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=21558421; PubMed=11702207;
 RA Kirschner R., Erturk D., Zeitz C., Sahin S., Ramser J.,
 RA Cremers F.P.M., Ropers H.-H., Berger W.;
 RA "DNA sequence comparison of human and mouse retinitis pigmentosa
 RT GTPase regulator (RPGR) identifies tissue-specific exons and putative
 RT regulatory elements.";

RL Hum. Genet. 109:271-278(2001).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Retina;
 RA Berger W.;
 RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SOURCE OF 525-840 FROM N.A. (ISOFORM 1), AND VARIANTS RP3 GLN-98 AND
 RP ARG-250.
 RX MEDLINE=20392404; PubMed=10932196;
 RA Verwoort R., Lennon A., Bird A.C., Tulloch B., Axton R., Milano M.G.,
 RA Meindl A., Mellinger T., Ciccodicola A., Wright A.F.;
 RT "Mutational hot spot within a new RPGR exon in X-linked retinitis
 RL pigmentosa.";
 RL Nat. Genet. 25:462-466(2000).
 RN [7]
 RP VARIANTS RP3 VAL-60 AND VAL-75, AND VARIANTS GLY-262; LYS-425; VAL-431
 RP AND GLY-566.
 RX MEDLINE=98153625; PubMed=9399904;
 RA Buraczynska M., Wu W., Fujita R., Buraczynska K., Phelps E.,
 RA Andreasson S., Bennett J., Birch D.G., Fishman G.A., Hoffman D.R.,
 RA Inana G., Jacobson S.G., Musarella M.A., Slevy P.A., Swarcop A.;
 RT "Spectrum of mutations in the RPGR gene that are identified in 20% of
 RL families with X-linked retinitis pigmentosa.";
 RL Am. J. Hum. Genet. 61:1287-1292(1997).
 RN [8]
 RP VARIANT RP3 VAL-60.
 RX MEDLINE=99070804; PubMed=9855162;
 RA Fishman G.A., Grover S., Jacobson S.G., Alexander K.R., Derlacki D.J.,
 RA Wu W., Buraczynska M., Swarcop A.;
 RT "X-linked retinitis pigmentosa in two families with a missense
 RT mutation in the RPGR gene and putative change of glycine to valine at
 RL codon 60.";
 RL Ophthalmology 105:2286-2296(1998).
 RN [9]
 RP VARIANTS RP3 ASN-99 AND VAL-289.
 RX MEDLINE=99415288; PubMed=10482958;
 RA Milano M.G., Testa F., Strazullo M., Trujillo M., De Bernardo C.,
 RA Grimalt B., Simonelli F., Mangino M., Torrente I., Ruberto G.,
 RA Beneyto M., Antinolo G., Rinaldi E., Danesino C., Venturto V.,
 RA D'Urso M., Ayuso C., Baiget M., Ciccodicola A.;
 RT "Mutation analysis of the RPGR gene reveals novel mutations in south
 RT European patients with X-linked retinitis pigmentosa.";
 RL Eur. J. Hum. Genet. 7:687-694(1999).
 RN [10]
 RP VARIANTS ILE-76; LYS-425 AND GLU-566.
 RX MEDLINE=99408233; PubMed=10480356;
 RA Zito I., Thiseleton D.L., Gorin M.B., Stout J.T., Plant C., Bird A.C.,
 RA Bhattacharya S.S., Hardcastle A.J.;
 RT "Identification of novel RPGR (retinitis pigmentosa GTPase regulator)
 RT mutations in a subset of X-linked retinitis pigmentosa families
 RL segregating with the RP3 locus.";
 RL Hum. Genet. 105:57-62(1999).
 RN [11]
 RP VARIANT RP3 ARG-302.
 RX MEDLINE=20202845; PubMed=10737996;
 RA Zito I., Gorin M.B., Plant C., Bird A.C., Bhattacharya S.S.,
 RA Hardcastle A.J.;
 RT "Novel mutations of the RPGR gene in RP3 families.";
 RL Hum. Mutat. 15:386-386(2000).
 RN [12]
 RP VARIANTS LYS-425; GLN-526 DEL; MET-533 AND GLU-566.
 RX MEDLINE=20438359; PubMed=10980513;
 RA Zito I., Morris A., Tyson P., Winship I., Sharp D., Gilbert D.,
 RA Thiseleton D.L., Bhattacharya S.S., Hardcastle A.J.;
 RT "Sequence variation within the RPGR gene: evidence for a founder
 RL complex allele.";
 RL Hum. Mutat. 16:273-274(2000).
 RN [13]
 RP VARIANT RP3 ASP-436.
 RX MEDLINE=21133224; PubMed=11180598;
 RA Guevara-Fujita M., Fahrner S., Buraczynska K., Cook J., Wheaton D.,
 RA Cortes F., Vicencio C., Pena M., Fishman G.A., Mintz-Hittner H.,

RA Birch D.G., Hoffman D.R., Mears A.J., Fujita R., Swarcop A.;
 RT "Five novel RPGR mutations in families with X-linked retinitis
 pigmentosa."
 RL Hum. Mutat. 17:151-151(2001).
 CC -1- FUNCTION: COULD BE GUANINE-NUCLEOTIDE RELEASING FACTOR.
 CC -1- SUBCELLULAR LOCATION: Golgi-associated (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing: Named isoforms-4;
 CC Comment-Additional isoforms seem to exist.
 CC Name-1;
 CC IsoId-Q92834-1: Sequence-Displayed;
 CC Name-2;
 CC IsoId-Q92834-2: Sequence-VSP_005548;
 CC Name-3;
 CC IsoId-Q92834-3: Sequence-VSP_005548, VSP_005549, VSP_005550;
 CC Name-4;
 CC IsoId-Q92834-4: Sequence-VSP_005547, VSP_005548;
 CC -1- TISSUE SPECIFICITY: Heart, brain, placenta, lung, liver, muscle,
 CC kidney, retina, pancreas and fetal retinal pigment epithelium.
 CC Isoform 3 is found only in the retina.
 CC -1- PTM: Prenylated (By similarity).
 CC -1- DISEASE: DEFECTS IN RPGR ARE THE CAUSE OF X-LINKED RETINITIS
 CC PIGMENTOSA-3 (XLRP-3 OR RP3), A FORM OF CHOROIDO-RETINAL
 CC DEGENERATION WHICH IS DISTINGUISHED FROM OTHER TYPES BY THE
 CC PRESENCE IN HETEROZYGOUS WOMEN OF A TAPETAL-LIKE RETINAL REFLEX (A
 CC BRILLIANT, SCINTILLATING, GOLDEN-HUED, PATCHY APPEARANCE MOST
 CC STRIKING AROUND THE MACULA) BUT NO VISUAL DEFECT.
 CC -1- SIMILARITY: Contains 6 RCC1 repeats.
 CC -1- DATABASE: NAME-Mutations of the RPGR gene;
 CC NOTE-beta International's Scientific Newsletter;
 CC WWW="http://www.retina-international.com/sci-news/rpgrmut.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U57629; AAC50481.1; -;
 DR EMBL: AJ238395; CAB54002.1; -;
 DR EMBL: AJ318463; CAC86116.1; -;
 DR EMBL: X97668; CAA66258.1; -;
 DR EMBL: AF286471; AAC00550.1; -;
 DR HSSP: P18754; 1A12.
 DR Genew: HGNC:10295; RPGR.
 DR MIM: 312610; -;
 DR MIM: 300389; -;
 DR GO: GO:0005794; C:Golgi apparatus; ISS.
 DR GO: GO:0019839; F:guanyl-nucleotide release factor activity; NAS.
 DR GO: GO:0005515; F:protein binding activity; TAS.
 DR GO: GO:0006886; P:intracellular protein transport; TAS.
 DR InterPro: IPR000408; Reg_chtr_condens.
 DR Pfam: PF00415; RCC1; 4.
 DR PROSITE: PS00625; RCC1_1; FALSE_NEG.
 DR PROSITE: PS00626; RCC1_2; 4.
 DR PROSITE: PS00626; RCC1_3; 6.
 DR PROSITE: PS00626; RCC1_4; 6.
 DR PROSITE: PS00626; RCC1_5; 6.
 DR PROSITE: PS00626; RCC1_6; 6.
 DR PROSITE: PS00626; RCC1_7; 6.
 DR PROSITE: PS00626; RCC1_8; 6.
 DR PROSITE: PS00626; RCC1_9; 6.
 DR PROSITE: PS00626; RCC1_10; 6.
 DR PROSITE: PS00626; RCC1_11; 6.
 DR PROSITE: PS00626; RCC1_12; 6.
 DR PROSITE: PS00626; RCC1_13; 6.
 DR PROSITE: PS00626; RCC1_14; 6.
 DR PROSITE: PS00626; RCC1_15; 6.
 DR PROSITE: PS00626; RCC1_16; 6.
 DR PROSITE: PS00626; RCC1_17; 6.
 DR PROSITE: PS00626; RCC1_18; 6.
 DR PROSITE: PS00626; RCC1_19; 6.
 DR PROSITE: PS00626; RCC1_20; 6.
 DR PROSITE: PS00626; RCC1_21; 6.
 DR PROSITE: PS00626; RCC1_22; 6.
 DR PROSITE: PS00626; RCC1_23; 6.
 DR PROSITE: PS00626; RCC1_24; 6.
 DR PROSITE: PS00626; RCC1_25; 6.
 DR PROSITE: PS00626; RCC1_26; 6.
 DR PROSITE: PS00626; RCC1_27; 6.
 DR PROSITE: PS00626; RCC1_28; 6.
 DR PROSITE: PS00626; RCC1_29; 6.
 DR PROSITE: PS00626; RCC1_30; 6.
 DR PROSITE: PS00626; RCC1_31; 6.
 DR PROSITE: PS00626; RCC1_32; 6.
 DR PROSITE: PS00626; RCC1_33; 6.
 DR PROSITE: PS00626; RCC1_34; 6.
 DR PROSITE: PS00626; RCC1_35; 6.
 DR PROSITE: PS00626; RCC1_36; 6.
 DR PROSITE: PS00626; RCC1_37; 6.
 DR PROSITE: PS00626; RCC1_38; 6.
 DR PROSITE: PS00626; RCC1_39; 6.
 DR PROSITE: PS00626; RCC1_40; 6.
 DR PROSITE: PS00626; RCC1_41; 6.
 DR PROSITE: PS00626; RCC1_42; 6.
 DR PROSITE: PS00626; RCC1_43; 6.
 DR PROSITE: PS00626; RCC1_44; 6.
 DR PROSITE: PS00626; RCC1_45; 6.
 DR PROSITE: PS00626; RCC1_46; 6.
 DR PROSITE: PS00626; RCC1_47; 6.
 DR PROSITE: PS00626; RCC1_48; 6.
 DR PROSITE: PS00626; RCC1_49; 6.
 DR PROSITE: PS00626; RCC1_50; 6.
 DR PROSITE: PS00626; RCC1_51; 6.
 DR PROSITE: PS00626; RCC1_52; 6.
 DR PROSITE: PS00626; RCC1_53; 6.
 DR PROSITE: PS00626; RCC1_54; 6.
 DR PROSITE: PS00626; RCC1_55; 6.
 DR PROSITE: PS00626; RCC1_56; 6.
 DR PROSITE: PS00626; RCC1_57; 6.
 DR PROSITE: PS00626; RCC1_58; 6.
 DR PROSITE: PS00626; RCC1_59; 6.
 DR PROSITE: PS00626; RCC1_60; 6.
 DR PROSITE: PS00626; RCC1_61; 6.
 DR PROSITE: PS00626; RCC1_62; 6.
 DR PROSITE: PS00626; RCC1_63; 6.
 DR PROSITE: PS00626; RCC1_64; 6.
 DR PROSITE: PS00626; RCC1_65; 6.
 DR PROSITE: PS00626; RCC1_66; 6.
 DR PROSITE: PS00626; RCC1_67; 6.
 DR PROSITE: PS00626; RCC1_68; 6.
 DR PROSITE: PS00626; RCC1_69; 6.
 DR PROSITE: PS00626; RCC1_70; 6.
 DR PROSITE: PS00626; RCC1_71; 6.
 DR PROSITE: PS00626; RCC1_72; 6.
 DR PROSITE: PS00626; RCC1_73; 6.
 DR PROSITE: PS00626; RCC1_74; 6.
 DR PROSITE: PS00626; RCC1_75; 6.
 DR PROSITE: PS00626; RCC1_76; 6.
 DR PROSITE: PS00626; RCC1_77; 6.
 DR PROSITE: PS00626; RCC1_78; 6.
 DR PROSITE: PS00626; RCC1_79; 6.
 DR PROSITE: PS00626; RCC1_80; 6.
 DR PROSITE: PS00626; RCC1_81; 6.
 DR PROSITE: PS00626; RCC1_82; 6.
 DR PROSITE: PS00626; RCC1_83; 6.
 DR PROSITE: PS00626; RCC1_84; 6.
 DR PROSITE: PS00626; RCC1_85; 6.
 DR PROSITE: PS00626; RCC1_86; 6.
 DR PROSITE: PS00626; RCC1_87; 6.
 DR PROSITE: PS00626; RCC1_88; 6.
 DR PROSITE: PS00626; RCC1_89; 6.
 DR PROSITE: PS00626; RCC1_90; 6.
 DR PROSITE: PS00626; RCC1_91; 6.
 DR PROSITE: PS00626; RCC1_92; 6.
 DR PROSITE: PS00626; RCC1_93; 6.
 DR PROSITE: PS00626; RCC1_94; 6.
 DR PROSITE: PS00626; RCC1_95; 6.
 DR PROSITE: PS00626; RCC1_96; 6.
 DR PROSITE: PS00626; RCC1_97; 6.
 DR PROSITE: PS00626; RCC1_98; 6.
 DR PROSITE: PS00626; RCC1_99; 6.
 DR PROSITE: PS00626; RCC1_100; 6.
 DR PROSITE: PS00626; RCC1_101; 6.
 DR PROSITE: PS00626; RCC1_102; 6.
 DR PROSITE: PS00626; RCC1_103; 6.
 DR PROSITE: PS00626; RCC1_104; 6.
 DR PROSITE: PS00626; RCC1_105; 6.
 DR PROSITE: PS00626; RCC1_106; 6.
 DR PROSITE: PS00626; RCC1_107; 6.
 DR PROSITE: PS00626; RCC1_108; 6.
 DR PROSITE: PS00626; RCC1_109; 6.
 DR PROSITE: PS00626; RCC1_110; 6.
 DR PROSITE: PS00626; RCC1_111; 6.
 DR PROSITE: PS00626; RCC1_112; 6.
 DR PROSITE: PS00626; RCC1_113; 6.
 DR PROSITE: PS00626; RCC1_114; 6.
 DR PROSITE: PS00626; RCC1_115; 6.
 DR PROSITE: PS00626; RCC1_116; 6.
 DR PROSITE: PS00626; RCC1_117; 6.
 DR PROSITE: PS00626; RCC1_118; 6.
 DR PROSITE: PS00626; RCC1_119; 6.
 DR PROSITE: PS00626; RCC1_120; 6.
 DR PROSITE: PS00626; RCC1_121; 6.
 DR PROSITE: PS00626; RCC1_122; 6.
 DR PROSITE: PS00626; RCC1_123; 6.
 DR PROSITE: PS00626; RCC1_124; 6.
 DR PROSITE: PS00626; RCC1_125; 6.
 DR PROSITE: PS00626; RCC1_126; 6.
 DR PROSITE: PS00626; RCC1_127; 6.
 DR PROSITE: PS00626; RCC1_128; 6.
 DR PROSITE: PS00626; RCC1_129; 6.
 DR PROSITE: PS00626; RCC1_130; 6.
 DR PROSITE: PS00626; RCC1_131; 6.
 DR PROSITE: PS00626; RCC1_132; 6.
 DR PROSITE: PS00626; RCC1_133; 6.
 DR PROSITE: PS00626; RCC1_134; 6.
 DR PROSITE: PS00626; RCC1_135; 6.
 DR PROSITE: PS00626; RCC1_136; 6.
 DR PROSITE: PS00626; RCC1_137; 6.
 DR PROSITE: PS00626; RCC1_138; 6.
 DR PROSITE: PS00626; RCC1_139; 6.
 DR PROSITE: PS00626; RCC1_140; 6.
 DR PROSITE: PS00626; RCC1_141; 6.
 DR PROSITE: PS00626; RCC1_142; 6.
 DR PROSITE: PS00626; RCC1_143; 6.
 DR PROSITE: PS00626; RCC1_144; 6.
 DR PROSITE: PS00626; RCC1_145; 6.
 DR PROSITE: PS00626; RCC1_146; 6.
 DR PROSITE: PS00626; RCC1_147; 6.
 DR PROSITE: PS00626; RCC1_148; 6.
 DR PROSITE: PS00626; RCC1_149; 6.
 DR PROSITE: PS00626; RCC1_150; 6.
 DR PROSITE: PS00626; RCC1_151; 6.
 DR PROSITE: PS00626; RCC1_152; 6.
 DR PROSITE: PS00626; RCC1_153; 6.
 DR PROSITE: PS00626; RCC1_154; 6.
 DR PROSITE: PS00626; RCC1_155; 6.
 DR PROSITE: PS00626; RCC1_156; 6.
 DR PROSITE: PS00626; RCC1_157; 6.
 DR PROSITE: PS00626; RCC1_158; 6.
 DR PROSITE: PS00626; RCC1_159; 6.
 DR PROSITE: PS00626; RCC1_160; 6.
 DR PROSITE: PS00626; RCC1_161; 6.
 DR PROSITE: PS00626; RCC1_162; 6.
 DR PROSITE: PS00626; RCC1_163; 6.
 DR PROSITE: PS00626; RCC1_164; 6.
 DR PROSITE: PS00626; RCC1_165; 6.
 DR PROSITE: PS00626; RCC1_166; 6.
 DR PROSITE: PS00626; RCC1_167; 6.
 DR PROSITE: PS00626; RCC1_168; 6.
 DR PROSITE: PS00626; RCC1_169; 6.
 DR PROSITE: PS00626; RCC1_170; 6.
 DR PROSITE: PS00626; RCC1_171; 6.
 DR PROSITE: PS00626; RCC1_172; 6.
 DR PROSITE: PS00626; RCC1_173; 6.
 DR PROSITE: PS00626; RCC1_174; 6.
 DR PROSITE: PS00626; RCC1_175; 6.
 DR PROSITE: PS00626; RCC1_176; 6.
 DR PROSITE: PS00626; RCC1_177; 6.
 DR PROSITE: PS00626; RCC1_178; 6.
 DR PROSITE: PS00626; RCC1_179; 6.
 DR PROSITE: PS00626; RCC1_180; 6.
 DR PROSITE: PS00626; RCC1_181; 6.
 DR PROSITE: PS00626; RCC1_182; 6.
 DR PROSITE: PS00626; RCC1_183; 6.
 DR PROSITE: PS00626; RCC1_184; 6.
 DR PROSITE: PS00626; RCC1_185; 6.
 DR PROSITE: PS00626; RCC1_186; 6.
 DR PROSITE: PS00626; RCC1_187; 6.
 DR PROSITE: PS00626; RCC1_188; 6.
 DR PROSITE: PS00626; RCC1_189; 6.
 DR PROSITE: PS00626; RCC1_190; 6.
 DR PROSITE: PS00626; RCC1_191; 6.
 DR PROSITE: PS00626; RCC1_192; 6.
 DR PROSITE: PS00626; RCC1_193; 6.
 DR PROSITE: PS00626; RCC1_194; 6.
 DR PROSITE: PS00626; RCC1_195; 6.
 DR PROSITE: PS00626; RCC1_196; 6.
 DR PROSITE: PS00626; RCC1_197; 6.
 DR PROSITE: PS00626; RCC1_198; 6.
 DR PROSITE: PS00626; RCC1_199; 6.
 DR PROSITE: PS00626; RCC1_200; 6.
 DR PROSITE: PS00626; RCC1_201; 6.
 DR PROSITE: PS00626; RCC1_202; 6.
 DR PROSITE: PS00626; RCC1_203; 6.
 DR PROSITE: PS00626; RCC1_204; 6.
 DR PROSITE: PS00626; RCC1_205; 6.
 DR PROSITE: PS00626; RCC1_206; 6.
 DR PROSITE: PS00626; RCC1_207; 6.
 DR PROSITE: PS00626; RCC1_208; 6.
 DR PROSITE: PS00626; RCC1_209; 6.
 DR PROSITE: PS00626; RCC1_210; 6.
 DR PROSITE: PS00626; RCC1_211; 6.
 DR PROSITE: PS00626; RCC1_212; 6.
 DR PROSITE: PS00626; RCC1_213; 6.
 DR PROSITE: PS00626; RCC1_214; 6.
 DR PROSITE: PS00626; RCC1_215; 6.
 DR PROSITE: PS00626; RCC1_216; 6.
 DR PROSITE: PS00626; RCC1_217; 6.
 DR PROSITE: PS00626; RCC1_218; 6.
 DR PROSITE: PS00626; RCC1_219; 6.
 DR PROSITE: PS00626; RCC1_220; 6.
 DR PROSITE: PS00626; RCC1_221; 6.
 DR PROSITE: PS00626; RCC1_222; 6.
 DR PROSITE: PS00626; RCC1_223; 6.
 DR PROSITE: PS00626; RCC1_224; 6.
 DR PROSITE: PS00626; RCC1_225; 6.
 DR PROSITE: PS00626; RCC1_226; 6.
 DR PROSITE: PS00626; RCC1_227; 6.
 DR PROSITE: PS00626; RCC1_228; 6.
 DR PROSITE: PS00626; RCC1_229; 6.
 DR PROSITE: PS00626; RCC1_230; 6.
 DR PROSITE: PS00626; RCC1_231; 6.
 DR PROSITE: PS00626; RCC1_232; 6.
 DR PROSITE: PS00626; RCC1_233; 6.
 DR PROSITE: PS00626; RCC1_234; 6.
 DR PROSITE: PS00626; RCC1_235; 6.
 DR PROSITE: PS00626; RCC1_236; 6.
 DR PROSITE: PS00626; RCC1_237; 6.
 DR PROSITE: PS00626; RCC1_238; 6.
 DR PROSITE: PS00626; RCC1_239; 6.
 DR PROSITE: PS00626; RCC1_240; 6.
 DR PROSITE: PS00626; RCC1_241; 6.
 DR PROSITE: PS00626; RCC1_242; 6.
 DR PROSITE: PS00626; RCC1_243; 6.
 DR PROSITE: PS00626; RCC1_244; 6.
 DR PROSITE: PS00626; RCC1_245; 6.
 DR PROSITE: PS00626; RCC1_246; 6.
 DR PROSITE: PS00626; RCC1_247; 6.
 DR PROSITE: PS00626; RCC1_248; 6.
 DR PROSITE: PS00626; RCC1_249; 6.
 DR PROSITE: PS00626; RCC1_250; 6.
 DR PROSITE: PS00626; RCC1_251; 6.
 DR PROSITE: PS00626; RCC1_252; 6.
 DR PROSITE: PS00626; RCC1_253; 6.
 DR PROSITE: PS00626; RCC1_254; 6.
 DR PROSITE: PS00626; RCC1_255; 6.
 DR PROSITE: PS00626; RCC1_256; 6.
 DR PROSITE: PS00626; RCC1_257; 6.
 DR PROSITE: PS00626; RCC1_258; 6.
 DR PROSITE: PS00626; RCC1_259; 6.
 DR PROSITE: PS00626; RCC1_260; 6.
 DR PROSITE: PS00626; RCC1_261; 6.
 DR PROSITE: PS00626; RCC1_262; 6.
 DR PROSITE: PS00626; RCC1_263; 6.
 DR PROSITE: PS00626; RCC1_264; 6.
 DR PROSITE: PS00626; RCC1_265; 6.
 DR PROSITE: PS00626; RCC1_266; 6.
 DR PROSITE: PS00626; RCC1_267; 6.
 DR PROSITE: PS00626; RCC1_268; 6.
 DR PROSITE: PS00626; RCC1_269; 6.
 DR PROSITE: PS00626; RCC1_270; 6.
 DR PROSITE: PS00626; RCC1_271; 6.
 DR PROSITE: PS00626; RCC1_272; 6.
 DR PROSITE: PS00626; RCC1_273; 6.
 DR PROSITE: PS00626; RCC1_274; 6.
 DR PROSITE: PS00626; RCC1_275; 6.
 DR PROSITE: PS00626; RCC1_276; 6.
 DR PROSITE: PS00626; RCC1_277; 6.
 DR PROSITE: PS00626; RCC1_278; 6.
 DR PROSITE: PS00626; RCC1_279; 6.
 DR PROSITE: PS00626; RCC1_280; 6.
 DR PROSITE: PS00626; RCC1_281; 6.
 DR PROSITE: PS00626; RCC1_282; 6.
 DR PROSITE: PS00626; RCC1_283; 6.
 DR PROSITE: PS00626; RCC1_284; 6.
 DR PROSITE: PS00626; RCC1_285; 6.
 DR PROSITE: PS00626; RCC1_286; 6.
 DR PROSITE: PS00626; RCC1_287; 6.
 DR PROSITE: PS00626; RCC1_288; 6.
 DR PROSITE: PS00626; RCC1_289; 6.
 DR PROSITE: PS00626; RCC1_290; 6.
 DR PROSITE: PS00626; RCC1_291; 6.
 DR PROSITE: PS00626; RCC1_292; 6.
 DR PROSITE: PS00626; RCC1_293; 6.
 DR PROSITE: PS00626; RCC1_294; 6.
 DR PROSITE: PS00626; RCC1_295; 6.
 DR PROSITE: PS00626; RCC1_296; 6.
 DR PROSITE: PS00626; RCC1_297; 6.
 DR PROSITE: PS00626; RCC1_298; 6.
 DR PROSITE: PS00626; RCC1_299; 6.
 DR PROSITE: PS00626; RCC1_300; 6.
 DR PROSITE: PS00626; RCC1_301; 6.
 DR PROSITE: PS00626; RCC1_302; 6.
 DR PROSITE: PS00626; RCC1_303; 6.
 DR PROSITE: PS00626; RCC1_304; 6.
 DR PROSITE: PS00626; RCC1_305; 6.
 DR PROSITE: PS00626; RCC1_306; 6.
 DR PROSITE: PS00626; RCC1_307; 6.
 DR PROSITE: PS00626; RCC1_308; 6.
 DR PROSITE: PS00626; RCC1_309; 6.
 DR PROSITE: PS00626; RCC1_310; 6.
 DR PROSITE: PS00626; RCC1_311; 6.
 DR PROSITE: PS00626; RCC1_312; 6.
 DR PROSITE: PS00626; RCC1_313; 6.
 DR PROSITE: PS00626; RCC1_314; 6.
 DR PROSITE: PS00626; RCC1_315; 6.
 DR PROSITE: PS00626; RCC1_316; 6.
 DR PROSITE: PS00626; RCC1_317; 6.
 DR PROSITE: PS00626; RCC1_318; 6.
 DR PROSITE: PS00626; RCC1_319; 6.
 DR PROSITE: PS00626; RCC1_320; 6.
 DR PROSITE: PS00626; RCC1_321; 6.
 DR PROSITE: PS00626; RCC1_322; 6.
 DR PROSITE: PS00626; RCC1_323; 6.
 DR PROSITE: PS00626; RCC1_324; 6.
 DR PROSITE: PS00626; RCC1_325; 6.
 DR PROSITE: PS00626; RCC1_326; 6.
 DR PROSITE: PS00626; RCC1_327; 6.
 DR PROSITE: PS00626; RCC1_328; 6.
 DR PROSITE: PS00626; RCC1_329; 6.
 DR PROSITE: PS00626; RCC1_330; 6.
 DR PROSITE: PS00626; RCC1_331; 6.
 DR PROSITE: PS00626; RCC1_332; 6.
 DR PROSITE: PS00626; RCC1_333; 6.
 DR PROSITE: PS00626; RCC1_334; 6.
 DR PROSITE: PS00626; RCC1_335; 6.
 DR PROSITE: PS00626; RCC1_336; 6.
 DR PROSITE: PS00626; RCC1_337; 6.
 DR PROSITE: PS00626; RCC1_338; 6.
 DR PROSITE: PS00626; RCC1_339; 6.
 DR PROSITE: PS00626; RCC1_340; 6.
 DR PROSITE: PS00626; RCC1_341; 6.
 DR PROSITE: PS00626; RCC1_342; 6.
 DR PROSITE: PS00626; RCC1_343; 6.
 DR PROSITE: PS00626; RCC1_344; 6.
 DR PROSITE: PS00626; RCC1_345; 6.
 DR PROSITE: PS00626; RCC1_346; 6.
 DR PROSITE: PS00626; RCC1_347; 6.
 DR PROSITE: PS00626; RCC1_348; 6.
 DR PROSITE: PS00626; RCC1_349; 6.
 DR PROSITE: PS00626; RCC1_350; 6.
 DR PROSITE: PS00626; RCC1_351; 6.
 DR PROSITE: PS00626; RCC1_352; 6.
 DR PROSITE: PS00626; RCC1_353; 6.
 DR PROSITE: PS00626; RCC1_354; 6.
 DR PROSITE: PS00626; RCC1_355; 6.
 DR PROSITE: PS00626; RCC1_356; 6.
 DR PROSITE: PS00626; RCC1_357; 6.
 DR PROSITE: PS00626; RCC1_358; 6.
 DR PROSITE: PS00626; RCC1_359; 6.
 DR PROSITE: PS00626; RCC1_360; 6.
 DR PROSITE: PS00626; RCC1_361; 6.
 DR PROSITE: PS00626; RCC1_362; 6.
 DR PROSITE: PS00626; RCC1_363; 6.
 DR PROSITE: PS00626; RCC1_364; 6.
 DR PROSITE: PS00626; RCC1_365; 6.
 DR PROSITE: PS00626; RCC1_366; 6.
 DR PROSITE: PS00626; RCC1_367; 6.
 DR PROSITE: PS00626; RCC1_368; 6.
 DR PROSITE: PS00626; RCC1_369; 6.
 DR PROSITE: PS00626; RCC1_370; 6.
 DR PROSITE: PS00626; RCC1_371; 6.
 DR PROSITE: PS00626; RCC1_372; 6.
 DR PROSITE: PS00626; RCC1_373; 6.
 DR PROSITE: PS00626; RCC1_374; 6.
 DR PROSITE: PS00626; RCC1_375; 6.
 DR PROSITE: PS00626; RCC1_376; 6.
 DR PROSITE: PS00626; RCC1_377; 6.
 DR PROSITE: PS00626; RCC1_378; 6.
 DR PROSITE: PS00626; RCC1_379; 6.
 DR PROSITE: PS00626; RCC1_380; 6.
 DR PROSITE: PS00626; RCC1_381; 6.
 DR PROSITE: PS00626; RCC1_382; 6.
 DR PROSITE: PS00626; RCC1_383; 6.
 DR PROSITE: PS00626; RCC1_384; 6.
 DR PROSITE: PS00626; RCC1_385; 6.
 DR PROSITE: PS00626; RCC1_386; 6.
 DR PROSITE: PS00626; RCC1_387; 6.
 DR PROSITE: PS00626; RCC1_388; 6.
 DR PROSITE: PS00626; RCC1_389; 6.
 DR PROSITE: PS00626; RCC1_390; 6.
 DR PROSITE: PS00626; RCC1_391; 6.
 DR PROSITE: PS00626; RCC1_392; 6.
 DR PROSITE: PS00626; RCC1_393; 6.
 DR PROSITE: PS00626; RCC1_394; 6.
 DR PROSITE: PS00626; RCC1_395; 6.
 DR PROSITE: PS00626; RCC1_396; 6.
 DR PROSITE: PS00626; RCC1_397; 6.
 DR PROSITE: PS00626; RCC1_398; 6.
 DR PROSITE: PS00626; RCC1_399; 6.
 DR PROSITE: PS00626; RCC1_400; 6.
 DR PROSITE: PS00626; RCC1_401; 6.
 DR PROSITE: PS00626; RCC1_402; 6.
 DR PROSITE: PS00626; RCC1_403; 6.
 DR PROSITE: PS00626; RCC1_404; 6.
 DR PROSITE: PS00626; RCC1_405; 6.
 DR PROSITE: PS00626; RCC1_406; 6.
 DR PROSITE: PS00626; RCC1_407; 6.
 DR PROSITE: PS00626; RCC1_408; 6.
 DR PROSITE: PS00626; RCC1_409; 6.
 DR PROSITE: PS00626; RCC1_410; 6.
 DR PROSITE: PS00626; RCC1_411; 6.
 DR PROSITE: PS00626; RCC1_412; 6.
 DR PROSITE: PS00626; RCC1_413; 6.
 DR PROSITE: PS00626; RCC1_414; 6.
 DR PROSITE: PS00626; RCC1_415; 6.
 DR PROSITE: PS00626; RCC1_416; 6.
 DR PROSITE: PS00626; RCC1_417; 6.
 DR PROSITE: PS00626; RCC1_418; 6.
 DR PROSITE: PS00626; RCC1_419; 6.
 DR PROSITE: PS00626; RCC1_420; 6.
 DR PROSITE: PS00626; RCC1_421; 6.
 DR PROSITE: PS00626; RCC1_422; 6.
 DR PROSITE: PS00626; RCC1_423; 6.
 DR PROSITE: PS00626; RCC1_424; 6.
 DR PROSITE: PS00626; RCC1_425; 6.
 DR PROSITE: PS00626; RCC1_426; 6.
 DR PROSITE: PS00626; RCC1_427; 6.
 DR PROSITE: PS00626; RCC1_428; 6.
 DR PROSITE: PS00626; RCC1_429; 6.
 DR PROSITE: PS00626; RCC1_430; 6.
 DR PROSITE: PS00626; RCC1_431; 6.
 DR PROSITE: PS00626; RCC1_432; 6.
 DR PROSITE: PS00626; RCC1_433; 6.
 DR PROSITE: PS00626; RCC1_434; 6.
 DR PROSITE: PS00626; RCC1_435; 6.
 DR PROSITE: PS00626; RCC1_436; 6.
 DR PROSITE: PS00626; RCC1_437; 6.
 DR PROSITE: PS00626; RCC1_438; 6.
 DR PROSITE: PS00626; RCC1_439; 6.
 DR PROSITE: PS00626; RCC1_440; 6.
 DR PROSITE: PS00626; RCC1_441; 6.
 DR PROSITE: PS00626; RCC1_442; 6.
 DR PROSITE: PS00626; RCC1_443; 6.
 DR PROSITE: PS00626; RCC1_444; 6.
 DR PROSITE: PS00626; RCC1_445; 6.
 DR PROSITE: PS00626; RCC1_446; 6.
 DR PROSITE:

